

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 12:46:11 ; Search time 7196 Seconds  
(without alignments)  
10938.038 Million cell updates/sec

Title: US-10-018-902-3

Perfect score: 1224

Sequence: 1 ccacgcctccgcacatcaataa.....aaaaaaaaaaaaaaaaaaag 1924

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pi.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_ats.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rnd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1924	100.0	1924	6	AX081073	AX081073 Sequence
2	1236.2	64.3	1703	8	AB011416	AB011416 Oryza sat
3	1216.4	63.2	1938	8	HVGTNRARI	X92403 H.vulgare m
4	1215.6	63.2	1997	8	HVHEMARI	X8101 H.vulgare m
5	1214.2	63.1	1864	8	D88382	D88382 Hordeum vul
6	1196.4	62.2	1778	6	AX081077	AX081077 Sequence
7	1107	57.5	1480	6	AX081089	AX081089 Sequence
8	1100.4	57.2	1822	8	D88383	D88383 Hordeum vul
9	1090.6	56.7	1703	8	HVHEMARI	X8101 H.vulgare m
10	987	51.3	130058	8	AC078840	AC078840 Oryza sat
11	987	51.3	300029	8	AE017109	AE017109 Oryza sat
12	710.4	36.9	2055	6	AX081081	AX081081 Sequence
13	706.2	36.7	2072	8	D50407	D50407 Cucurbit m
14	679.6	35.3	1632	6	AX081085	AX081085 Sequence
15	679.6	35.3	1663	8	AY096600	AY096600 Arabidops
16	679.6	35.3	2111	8	AY072223	AY072223 Arabidops
17	619.8	32.2	1593	6	AX651934	AX651934 Sequence
18	602	31.3	1901	8	D67088	D67088 Cucurbit m
19	589.2	30.6	1983	6	AX081085	AX081085 Sequence
20	558.6	29.0	3663	8	AF105221	AF105221 Glycine m
21	541.2	28.1	86014	8	FL9C14	FL9C14 Sequence
22	534.6	27.8	2879	8	ATU03774	ATU03774 Arabidops
23	509.6	26.5	1979	8	AF294753	AF294753 Hordeum v
24	482.2	25.1	118803	8	AC124214	AC124214 Medicago
25	475.6	24.7	1047	8	AB075573	AB075573 Glyptost
26	472.4	24.6	1047	8	AB075574	AB075574 Taxodium
27	467	24.3	3683	8	AB075575	AB075575 Metasequo
28	467	24.3	3683	8	ATU27118	ATU27118 Arabidops
29	467	24.3	146505	8	F21M12	F21M12 Sequence
30	464.4	24.1	1047	8	AB075580	AB075580 Juniperus
31	464.4	24.1	1047	8	AB075576	AB075576 Sequoia s
32	464.4	24.1	1047	8	AB075582	AB075582 Cryptomer
33	461.2	24.0	1047	8	AB075577	AB075577 Chamaecyp
34	461.2	24.0	1047	8	AB075579	AB075579 Thuja sta
35	461.2	24.0	1047	8	AB075581	AB075581 Chamaecyp
36	458	23.8	1047	8	AB075578	AB075578 Thujaopsi
37	444	23.1	3179	8	AF305613	AF305613 Chlamydom
38	380.6	19.8	84872	8	AC006593	AC006593 Arabidops
39	329	17.1	344	8	AF496730	AF496730 Zea mays
40	329	17.1	344	8	AF496731	AF496731 Zea mays
41	329	17.1	344	8	AF496732	AF496732 Zea mays
42	329	17.1	344	8	AF496733	AF496733 Zea mays
43	329	17.1	344	8	AF496734	AF496734 Zea mays
44	329	17.1	344	8	AF496735	AF496735 Zea mays
45	329	17.1	344	8	AF496736	AF496736 Zea mays

# ALIGNMENTS

RESULT 1  
AX081073  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AX081073  
Sequence 3 from Patent WO0109304.  
AX081073  
AX081073.1 GI:13170000

1924 bp  
DNA  
linear  
PAT 27-FEB-2001

Zea mays  
Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
AUTHORS

1  
Cahoon, R.E., Gutteridge, S., Harvell, L.T., Rafalski, J.A., Tao, Y. and  
Weng, Z.

TITLE		Polynucleotides encoding aminolevulinic acid biosynthetic enzymes	
JOURNAL		Patent: WO 0109304-A 3 08-FEB-2001;	
FEATURES		E.I. DU PONT DE NEMOURS AND COMPANY (US)	
source	Location/Qualifiers	1..1924	
	/organism="Zea mays"		
	/mol_type="genomic DNA"		
BASE COUNT		504 a	465 c 551 g 404 t
ORIGIN			
Query Match		100.0%;	Score 1924; -DB 6; Length 1924;
Best Local Similarity		100.0%;	Pred. No. 0;
Matches 1924;		Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	CCACGCGTCCGATCAATAAGAGAGAGCTTGGGAAGTTGCCAAGGCCCTCCGATTCGCTA	60
Db	1	CCACGCGTCCGATCAATAAGAGAGAGCTTGGGAAGTTGCCAAGGCCCTCCGATTCGCTA	60
Qy	61	ATGCCACGATATGCGGACCGACGAGCTCAGCGACCAACCGCCGCGCAGCAGCCGACCA	120
Db	61	ATGCCACGATATGCGGACCGACGAGCTCAGCGACCAACCGCCGCGCAGCAGCCGACCA	120
Qy	121	CGCCAAAGCCGCGGGGCTGCTGCTCGGCCCTCTGCCAGAGGCTGGCCGCGCGCAGGC	180
Db	121	CGCCAAAGCCGCGGGGCTGCTGCTCGGCCCTCTGCCAGAGGCTGGCCGCGCGCAGGC	180
Qy	181	GGCGTCCGGGCTGGTGGCTGCGACGCGCGCGGCTGGAGGCCCGAGCGCGCTGG	240
Db	181	GGCGTCCGGGCTGGTGGCTGCGACGCGCGCGGCTGGAGGCCCGAGCGCGCTGG	240
Qy	241	CAAGGCTGCCAGCGCTCGCGCCCTCGAGCAGTTCAGATATCCGCGCCGCTACATGA	300
Db	241	CAAGGCTGCCAGCGCTCGCGCCCTCGAGCAGTTCAGATATCCGCGCCGCTACATGA	300
Qy	301	AGGAAGAGGATACCATAGCTGTGATAGGCTCAGTGTACACACAGCACCAAGTGGAGATGC	360
Db	301	AGGAAGAGGATACCATAGCTGTGATAGGCTCAGTGTACACACAGCACCAAGTGGAGATGC	360
Qy	361	GTGAATAACTTCTGTTGAGAGGAACCTGTGGCCCGCTGCTATTCAAGAACTCACTAGCC	420
Db	361	GTGAATAACTTCTGTTGAGAGGAACCTGTGGCCCGCTGCTATTCAAGAACTCACTAGCC	420
Qy	421	TGAACCATATTGAAGAGGCTGCTGTTCTTAGTACCTGTAATAGATGGAATTTATGCG	480
Db	421	TGAACCATATTGAAGAGGCTGCTGTTCTTAGTACCTGTAATAGATGGAATTTATGCG	480
Qy	481	TGGCGCTATCATGGAACCGTGTATCAGAGAAGTAGTGGACTGGATGTCGAAGAAAGTG	540
Db	481	TGGCGCTATCATGGAACCGTGTATCAGAGAAGTAGTGGACTGGATGTCGAAGAAAGTG	540
Qy	541	GTATTCGCGCTTCGAGCTTAGGGAGCACCTGTTTCATCTTTCGAGAGCAGTATGCCAC	600
Db	541	GTATTCGCGCTTCGAGCTTAGGGAGCACCTGTTTCATCTTTCGAGAGCAGTATGCCAC	600
Qy	601	GCCATCTGTTGAGGTGTGAGTGGCTTGAATCTTTGTTCTCGGTGAAGGACAAATCC	660
Db	601	GCCATCTGTTGAGGTGTGAGTGGCTTGAATCTTTGTTCTCGGTGAAGGACAAATCC	660
Qy	661	TTCCTCAGGTTAAACAAGTTGTGAGGAGTGGACAGAACAGTGGAGGCTTGGGAAGAACA	720
Db	661	TTCCTCAGGTTAAACAAGTTGTGAGGAGTGGACAGAACAGTGGAGGCTTGGGAAGAACA	720
Qy	721	TGATAGGATGTTCAAGATGCAATCACTCTGCTGGAAGAGCGTGTCCGACGAGCACAACA	780
Db	721	TGATAGGATGTTCAAGATGCAATCACTCTGCTGGAAGAGCGTGTCCGACGAGCACAACA	780
Qy	781	TATCATCTGGTGTCTTTCTGTGAGTTCAGCGCGGTTGAACTGGCCCTGATGAAGCTTC	840
Db	781	TATCATCTGGTGTCTTTCTGTGAGTTCAGCGCGGTTGAACTGGCCCTGATGAAGCTTC	840
Qy	841	CGAAGTCTGAAGCACTGTCAAGTAGGATGCTTCTGATTTGGTGGTAAATGGGAAGC	900
Db			
841		CGAAGTCTGAAGCACTGTCAAGTAGGATGCTTCTGATTTGGTGGTAAATGGGAAGC	900
Qy	901	TAGTGATCAAAACATCTGTTGCCAAAAGATGCAAGAGGTTGTTGTGTGAACCCCTCG	960
Db	901	TAGTGATCAAAACATCTGTTGCCAAAAGATGCAAGAGGTTGTTGTGTGAACCCCTCG	960
Qy	961	TGGAAGGTCGATGCTATTCTCGTGGAGAGATGAAGATATAGAGATCTGTGTACAGGCTC	1020
Db	961	TGGAAGGTCGATGCTATTCTCGTGGAGAGATGAAGATATAGAGATCTGTGTACAGGCTC	1020
Qy	1021	TCTCAGACATGTATCAAGCTGCTGAAGCTGATGTGTTCAACAGCACCCGATCTG	1080
Db	1021	TCTCAGACATGTATCAAGCTGCTGAAGCTGATGTGTTCAACAGCACCCGATCTG	1080
Qy	1081	AAACTTCATTTTTCGAAAAGAACACGACAGGACATCCCCCCTGTCTCTGATATATGG	1140
Db	1081	AAACTTCATTTTTCGAAAAGAACACGACAGGACATCCCCCCTGTCTCTGATATATGG	1140
Qy	1141	GAGGTGTTTCGCTGTTTGTGACATATCTGTCCACAGGAATGTGAGCGCATGTGTCTG	1200
Db	1141	GAGGTGTTTCGCTGTTTGTGACATATCTGTCCACAGGAATGTGAGCGCATGTGTCTG	1200
Qy	1201	AAGTTGGCGCTGCGAGTGTACAAATGTGACGAGCTTGAAGAGGTTGGTGAAGGCAACA	1260
Db	1201	AAGTTGGCGCTGCGAGTGTACAAATGTGACGAGCTTGAAGAGGTTGGTGAAGGCAACA	1260
Qy	1261	AGGAGACCCGCTCAGGAAAGCAATGGAGGCGCAGACAAATCATCCGAAAGACTGAGAC	1320
Db	1261	AGGAGACCCGCTCAGGAAAGCAATGGAGGCGCAGACAAATCATCATCCGAAAGACTGAGAC	1320
Qy	1321	GGTTTCGAGGATGAGGAGCTCGCTGGAGACCGTTCCAGCAATCAAGAAAGCTGAGGTCT	1380
Db	1321	GGTTTCGAGGATGAGGAGCTCGCTGGAGACCGTTCCAGCAATCAAGAAAGCTGAGGTCT	1380
Qy	1381	ACGCGGACAGGATCAGGGCTCGGAGCTCGAGAGAGTGCCTGCAGAAAGTAGTGTAGGACG	1440
Db	1381	ACGCGGACAGGATCAGGGCTCGGAGCTCGAGAGAGTGCCTGCAGAAAGTAGTGTAGGACG	1440
Qy	1441	CCCTCACCAAGAGATGAGGAGCCATCGAGGAGCTGAGCACCGGCATCTGTTAAACAAGC	1500
Db	1441	CCCTCACCAAGAGATGAGGAGCCATCGAGGAGCTGAGCACCGGCATCTGTTAAACAAGC	1500
Qy	1501	TCCTCCATGGCCCGCTGCAGCACCTGAGGTGCGAGCGGAGCAGCAGCCCTTTGAGC	1560
Db	1501	TCCTCCATGGCCCGCTGCAGCACCTGAGGTGCGAGCGGAGCAGCAGCCCTTTGAGC	1560
Qy	1561	AGACGCTCGAGAACATGCACGCCCTCAACCGGATGTTTCAGCCTCGACATGGAGAAAGCGA	1620
Db	1561	AGACGCTCGAGAACATGCACGCCCTCAACCGGATGTTTCAGCCTCGACATGGAGAAAGCGA	1620
Qy	1621	TCATCGAGCAGAGATCAAGGCCAAGTGGAGAGAGACACAAACTGAGGCGCAGGAAGCNA	1680
Db	1621	TCATCGAGCAGAGATCAAGGCCAAGTGGAGAGAGACACAAACTGAGGCGCAGGAAGCNA	1680
Qy	1681	TTTTTCTACCACTTATCTATATATATAGCGTCTCCAATCTCAATCTTTTATCC	1740
Db	1681	TTTTTCTACCACTTATCTATATATATAGCGTCTCCAATCTCAATCTTTTATCC	1740
Qy	1741	TTTCACTCAGTGTAGGCTTCCCTGCTCCTGCTGATCGTTAACTGTCTGTGAATTAGA	1800
Db	1741	TTTCACTCAGTGTAGGCTTCCCTGCTCCTGCTGATCGTTAACTGTCTGTGAATTAGA	1800
Qy	1801	GCCATGGCAGGCTGTTGTCAATAACAGCAATGTGTCCCAATTCGCCACAGAAAGAACT	1860
Db	1801	GCCATGGCAGGCTGTTGTCAATAACAGCAATGTGTCCCAATTCGCCACAGAAAGAACT	1860
Qy	1861	ATATTTATATGCAATTTATGGAGCAAAATAGTTTAAATAAAAAAATAAAAAA	1920
Db	1861	ATATTTATATGCAATTTATGGAGCAAAATAGTTTAAATAAAAAAATAAAAAA	1920
Qy	1921	AAAG 1924	
Db	1921	AAAG 1924	

```

RESULT 2
AB011416
LOCUS
DEFINITION
Oryza sativa (indica cultivar-group) mRNA for glutamyl-tRNA
reductase, complete cds.
ACCESSION
AB011416
VERSION
AB011416.1
KEYWORDS
glutamyl-tRNA reductase;
SOURCE
Oryza sativa (indica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1
AUTHORS
Nakayashiki, T. and Inokuchi, H.
TITLE
Nucleotide sequence of a cDNA clone encoding glutamyl-tRNA reductase
from rice (accession No. AB011416) (PCR 98-080)
JOURNAL
Plant Physiol. 117, 332 (1998)
REFERENCE
2
AUTHORS
Nakayashiki, T.
TITLE
Direct Submission
JOURNAL
Submitted (23-FEB-1998) Toru Nakayashiki, Graduate School of
Science, Kyoto University, Department of Biophysics, Sakyo-ku,
Kitashirakawa, Oiwakechou, Kyoto 606-8502, Japan
(E-mail: nakkun@molbio.biophys.kyoto-u.ac.jp, Tel: 075-753-4201)
FEATURES
Location/Qualifiers
1..1703
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/dev_stage="5-day-old"
72..1682
/codon_start=1
product="glutamyl-tRNA reductase"
/protein_id="BAA25003.1"
/db_xref="GI:2920320"
/translat="NMASSTSATAGGAAFAAKTRVVGGRRPCARVAAGRRRRSGV
VVRCDAGVEAQQAQVAAKAAASVALEQFKISADRYMKERSIAVIGISVHTAPVEMR
EKLAVALSELPSISLNLHIEEAAVLSTCNMEIYVVALSNRGLREYVDWMSKK
SGIPASLEHFLDRDSTLHLFEVSGALDSLVLGEIEILAOVKQVVRGQNSGGL
GKNDIMFKDIAITAGKVRCTNISSGAVSVSSAAVELAKMLPKSECLSRMLLIGA
GKMGKLVVKKHIAKCKVVVNRVERVDALREEMKDIEIVRPLTWMYEAAREADV
VFTSPASLFTKEHALPALSDAMGVRFLFVDISVPRNVSAVSVEGHARVTND
DLKEVVEANKDRLKAMEAQITITQELKRFEMRDSLETPTTIKKRSYADRIASE
LEKCLQKIGEDALTKKRRSIEELSTGIWNKLLHGLPLQHLRCDGSDSLDTLELNMH
ALNRFSLDTEKATIEQKIKAKVEKSN"
BASE COUNT
416 a 390 c 538 g 359 t
ORIGIN
Query Match 64.3%; Score 1236.2; DB 8; Length 1703;
Best Local Similarity 86.2%; Pred. No. 1.8e-202;
Matches 1406; Conservative 0; Mismatches 213; Indels 12; Gaps 3;
QY 68 GATATGCGGACACGACGTCAGCACACACGCGCGCGCAGCAGCGCCACACCGCCAA 127
DB |||||
DB 71 GATGATGCGGACGACGTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 130
QY 128 GCGCGCGGG- GTCGTCGTCGCGCCCTCTGCCA-----GAGGGTGGCGCGCGCGCGCGG 181
DB |||||
DB 131 GCGGCGGGTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 190
QY 182 GCGTCCGGGTGGTGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 235
DB |||||
DB 191 GCGGTCCGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 250
QY 236 CGTGGCAAGAGCTGCACGCTCGCGCGCGCTCGAGCAGTTCAAGATATCCGCGCGCGGTA 295
DB |||||
DB 251 GGTGCCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTA 310
QY 296 CATGAAGAAAGAGGTACCATAGTGTGTATAGGCGCTCAGTGTACACACAGCACCGAGTGA 355

```

```

DB |||||
QY 311 CATGAAGAAAGAGTAGCATAGCGGTAAATAGCGCTCAGTGTACACACTGCACCACTGGA 370
QY 356 GATCGGTGAAAAAATTGCTGTTGTCAGAGGAACCTGTGGCCCCGTGCTATTCAAGAACTCAC 415
DB |||||
DB 371 GATCGGTGAGAAACTTGTGTTGTCAGAGGAACCTATGGCCCCGTGCTATCTCAGAACTCAC 430
QY 416 TAGCCGTGAACCATATTGAAGAGGCTGCTGTTCTTAGTACCTGTAAATAGAAATTTA 475
DB |||||
DB 431 CAGTCTGAATCATATTGAAGAGGCTGCTGTTCTTAGTACCTGCAATAGAAATGAAATCTA 490
QY 476 TGTGTTGGCGCTCATCATGGAAACCGTGTATCAGAGAAGTAGTGGACTGATGTCGAAGAA 535
DB |||||
DB 491 TGTGTTAGCTTATCTGTTGAACCGTGGACTTAGAGAGTGGTAGACTGATGTCAAGAA 550
QY 536 AAGTGGTATTCCTGCTTCGAGCTTAGGGAGCACCTGTTTCATCTTCGGAAGCAGTGATGC 595
DB |||||
DB 551 AAGTGGAAATCCCTGCTTCTGAGCTCAGGGAGCATCTATTCAATGTTGCGTGACAGTGATGC 610
QY 596 CACAGCCCATCTGTTGAGGTGTCACTGGCCCTTCACTCTTTGGTCTCGGTGAAGGACA 655
DB |||||
DB 611 CACAGCCCATCTGTTTGGAGTATCTGCTGGGCTTCACTCTTTGGTCTTTGGAGAGCGCA 670
QY 656 AATCCTGCTCAGGTTAAACAAGTTGTGAGGAGTGGACAGAACAGTGGAGGCTTTGGAAA 715
DB |||||
DB 671 AATCCTGCTCAAGTTAAACAAGTTGTGAGAGTGGGCAACAGTGGAGGCTTTGGAAA 730
QY 716 GAACATCATAGGATGTTTCAAGGATGCAATCACTGCTCGGAAAGCGTGTCCGACGAGAC 775
DB |||||
DB 731 GAACATCATAGGATGTTTCAAGGATGCAATCACTGCTGGAAAGCGTGTCCGCTGCGAGAC 790
QY 776 CAACATATCATCTGCTGCTGTTCTGTCAGTTCAGCGGGGTTGAACTGGCCCTCATGAA 835
DB |||||
DB 791 TAAACATATCATCAGGTGCTGCTCTCTGTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 850
QY 836 GCTTCCGAAGTCTGAAGCACTGTCACTAGTAGATGCTTCTGATGTTGCTGTGTTGTTGTT 895
DB |||||
DB 851 GCTTCCAAGTTCGGAATGCCATCTGCTAGGATGCTGTTGTTGTTGTTGTTGTTGTTGTTG 910
QY 896 AAGCTAGTGTATCAACATCTGTTGTCGCAAGGATGCAAGAGGTTGTTGTTGTTGTTGTTG 955
DB |||||
DB 911 AAGTTGTTGTTTAAACATTTGATTGTCGCAAGGATGCAAGAGGTTGTTGTTGTTGTTGTT 970
QY 956 CTCGTTGAAAGGTTGGATGCTATTCGTGAGGAGATGAAAGATATAGAGATCGTGTACAG 1015
DB |||||
DB 971 TTCAGTGGAAAGGTTGGATGCTCATCCGCAAGAGATGAAAGACATTTGAGATTGTTGTTAC 1030
QY 1016 GCCTCTCTCAGACATGATCAAGCTGCTGCTGAAGCTGATGTCGTTGTTGTTGTTGTTGTTG 1075
DB |||||
DB 1031 GCCTCTTACAGAGATGATGAAGCGCTGCGAAGCTGATGTCGTTGTTGTTGTTGTTGTTGTT 1090
QY 1076 ATCTGAAACTTTCATTTGTTGCAAAAGAAACACGACAGAGGACATCCCGCTGTTCTGTATAC 1135
DB |||||
DB 1091 ATCCGAAACCCCATTTGTTTCAAAAGGACACGACAGAGGCGCTTCCCGCTATTTCTGATGC 1150
QY 1136 TATGGAGGTTTGGCTGCTGTTTGTGCAATATCTGTGTCGCGGAAATGTGACGCGCATGTGT 1195
DB |||||
DB 1151 TATGGGTGGTTCGACTCTTTGTGCAATATCCGTCGCGGAAATGTGACGCGCTGTTGT 1210
QY 1196 GTCTGAAAGTTCGGCTGCGACGAGTGTCAATGTGCGACACTTTGAAAGAGTGTGTTGAAAGC 1255
DB |||||
DB 1211 GTCTGAAAGTTCGGCTGCGACGAGTGTCAATGTGCGACACTTTGAAAGAGTGTGTTGAAAGC 1270
QY 1256 CAAACAAGAGGACCGGCTCAGGAAAGCAATGGAGGCGCAGACCAATCATCATCCGAGAAACT 1315
DB |||||
DB 1271 CAAACAAGAGGACCGGCTTAGGAAAGCAATGGAGGCGCAGACCAATCATCATCCGAGAAAT 1330
QY 1316 GAGACGTTTCAGGACATGGAGGACTCGTGGAGACCGTTCCGACCATCAAGAAAGCTGAG 1375
DB |||||
DB 1331 GAAACGTTTCAGGACATGGAGGACTCGTGGAGACCGTTCCGACCATCAAGAAAGCTGAG 1390
QY 1376 GTCTGACGCGACAGGATCAGGCGCTCGGAGCTCGAGAGTGCCTGCGAGAAAGTAGGTGA 1435

```

Db 1391 GTCTACGCGACAGGATCAGGGCTTCGGAGCTTGAGAGTGCCTCCAGAGATCGGCGA 1450  
QY 1436 GGACGCGCTTACCAAGAGATGAGGAGACCCATCGAGGAGCTGAGCACCGGATCGTTAA 1495  
Db 1451 AGACGCGCTTACCAAGAGACGAGAGATCCATCGAGGAGCTCAGCACCGGATCGTAA 1510  
QY 1496 CAAGCTCCCTCAATGGCGCGCTCGACGACCTGAGGTGCGACGCGGACGCGCACCT 1555  
Db 1511 CAAGCTTCTCCACGGGCCATTGACGACCTGAGATGTGACGCGGACGCGCACCT 1570  
QY 1556 TGACGAGCGCTCGAAGACATGACGCGCTTCAACCGGATGTTTCAAGCTCGACATGAGAA 1615  
Db 1571 CGATGAGCGCTGGAGAACATGACGCGCTTCAACGAGATGTTTCAAGCTCGACATGAGAA 1630  
QY 1616 GCGGATCATGACGAGAGATCAAGCCCAAGTGGGAGAGACACAAACTGAGGCCAGGA 1675  
Db 1631 GCGGATCATGAGAGAGATCAAGCCCAAGTGGGAGAGATCCAGAACTGAGATTGAAG 1690  
QY 1676 AGCAATTTTTC 1686  
Db 1691 AGAGATTTTTC 1701

## RESULT 3

HVGTRNAR1  
LOCUS HVGTRNAR1 1938 bp mRNA linear PLN 08-NOV-1996  
DEFINITION H.vulgare mRNA for glutamyl-tRNA reductase, 1st isoform.  
ACCESSION X92403  
VERSION X92403.1 GI:1041424  
KEYWORDS glutamyl-tRNA reductase; hema 1 gene; isoform 1.  
SOURCE Hordeum vulgare subsp. vulgare  
ORGANISM Hordeum vulgare subsp. vulgare  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
1  
Bougri,O. and Grimm,B.  
Members of a low-copy number gene family encoding glutamyl-tRNA  
reductase are differentially expressed in barley  
Plant J. 9 (6), 867-878 (1996)  
96267576  
8696365  
2 (bases 1 to 1938)  
Bougri,O.  
Direct Submission  
Submitted (16-OCT-1995) O. Bougri, Institute of Plant Genetics &  
Crop, Plant Research, Corrensstr. 3, 06466 Gatersleben, FRG  
Location/Qualifiers  
1. .1938  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/sub\_species="klages"  
/db\_xref="taxon:112509"  
/chromosome="5"  
/clone\_lib="bhal"  
/clone\_lib="lamba gtl1"  
/dev\_stage="12 d old seedlings"  
1. .1938  
/gene="hema1"  
1. .1938  
/evidence="experimental"  
61. 1644  
/gene="hema1"  
/note="isoform I"  
/codon\_start=1  
/product="glutamyl-tRNA reductase"  
/protein\_id="CAA63140.1"  
/db\_xref="GI:1041425"  
/translation="MAGTSATAAGAFAPAAKARGPAAACPWLVAAAGRRRSGVVRCD  
AGDQAASKASITIALSQFKISADRYMKSSIAVIGLSVHTAPVNRKLAVAL  
WPRATISLTHIEAAVLSTCNRMETIYVVALSNRGIREVVDWMSKKGPASELR

## FEATURES

source  
1. .1938  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/sub\_species="klages"  
/db\_xref="taxon:112509"  
/chromosome="5"  
/clone\_lib="bhal"  
/clone\_lib="lamba gtl1"  
/dev\_stage="12 d old seedlings"  
1. .1938  
/gene="hema1"  
1. .1938  
/evidence="experimental"  
61. 1644  
/gene="hema1"  
/note="isoform I"  
/codon\_start=1  
/product="glutamyl-tRNA reductase"  
/protein\_id="CAA63140.1"  
/db\_xref="GI:1041425"  
/translation="MAGTSATAAGAFAPAAKARGPAAACPWLVAAAGRRRSGVVRCD  
AGDQAASKASITIALSQFKISADRYMKSSIAVIGLSVHTAPVNRKLAVAL  
WPRATISLTHIEAAVLSTCNRMETIYVVALSNRGIREVVDWMSKKGPASELR

## gene

## mRNA

## CDS

EHLPMLRSDATRHLPFVSAGLSLVLEGQILAQVKQVVRNQNSGGIGKNIDRMFK  
DAITAGKRVRETNISAGVSSAAVELAMMKPKSECLSMILLIGAGKMKLVVK  
HLIAKCKKVVVRNRSVERDAIRREMDIIVYRPLTEMYEADADAVFTSTASES  
LLFTKEHAEVLPPISLAMGVRFLFVDISVPRNVGACLSEVHARVYVNDLKEVVEAN  
KEDRVKAMBAQTIITBLKRFPEARWDSLETPTIKLRSVADRIRASLEKLOKIG  
EDNLNKKRRSIEELSTGIVNKLHLGHLQHLRCDGSDSRSTLDETLENHMLNRMFSLD  
TEKAVLEQKIKAKVEKTS"  
61. .201  
/gene="hema1"  
/note="transit-peptide site"  
202. .1641  
/gene="hema1"  
/product="unnamed"  
BASE COUNT 494 a 447 c 559 g 438 t  
ORIGIN  
Query Match 63.2%; Score 1216.4; DB 8; Length 1938;  
Best Local Similarity 83.1%; Pred. No. 4.6e-199;  
Matches 1426; Conservative 0; Mismatches 276; Indels 14; Gaps 3;  
QY 53 TTTTCCTAATGCGACGATAATGGCGACCAACGACGTGACGACCAACCGCGCGCAGCAGC 112  
Db 42 TTTTGTGCTGCTGTGATGATGGCGGAGCGACGTGAG---CCACGGCAGCGCGCGCGC 98  
QY 113 CGCCACCAACCGCCCAAGCGCGGGGTGCTGCTGGGCCCTCTGCCAGAGGGTGGCGCGCG 172  
Db 99 ATTGCCCGCGCCAAAGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 158  
QY 173 CGCAGGCGCGCGCTCCGGGTGCTGCGGTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCA 232  
Db 159 CGGCGGAGGCGGTCCGGCGTCTGCGGTGCGACCGCGCGCGCGCGG-----ATGCACA 209  
QY 233 GGCCTGCGAAAGGTGCGCGGTGCGCGCTCGAGCAGTTCAAGATATCGCGCGCGCGCG 292  
Db 210 GCGCGCTGCAAGCGCGCGCGCATCACCGCGCTCGAGCAGTTCAAGATCTCGCGCAGCG 269  
QY 293 GTACATGAAGAAAGAGTACCATAGCTGTGATAGGCGCTCAGGTGACACACAGCAGCAGCT 352  
Db 270 GTACATGAAGAAAGAGTACCATAGCTGTGATAGGCGCTCAGGTGACACACAGCAGCAGCT 329  
QY 353 GGAGATGCTGAAAACTTGTCTGTGACAGAGAACTGTGGCGCGCGCGCGCGCGCGCGCGCG 412  
Db 330 GGAGATGCGGAAAACTTGTCTGTGACAGAGAACTGTGGCGCGCGCGCGCGCGCGCGCG 389  
QY 413 CACTAGCCTGAACCATATGAAAGGTGCTGTTCTTAGTACCTGTAATAGAAATGAAAT 472  
Db 390 CACCACTGTAATCATATTTGAAGAGGTGCTGTTCTTAGTACCTGCAACAGAAATGAAAT 449  
QY 473 TTATGCTGGCGCTCATCATGAAACCGTGTATCAGAGAGTAGTGGACTGATGTCGAA 532  
Db 450 ATATGCTGGCTTTTATCGTGAACCGTGTATTTAGAGAGTAGTAGTGGATGTCAA 509  
QY 533 GAAAGTGTATTTCCCGCTTCGAGCTTACGAGAGCACTGTTCATCTTCCGAGCAGTGA 592  
Db 510 GAAAGTGTATTTCCCGCTTCGAGCTTACGAGAGCACTGTTCATCTTCCGAGCAGTGA 569  
QY 593 TGCCACACGCGCATCTGTTTGAAGGTGCTCAGCTGGCGCTTCACTCTTTGGTCTCGTAAAG 652  
Db 570 TGCCACACGCGCATCTGTTTGAAGGTGCTCAGCTGGCGCTTCACTCTTTGGTCTCGTAAAG 629  
QY 653 ACAATCTCTCTCAGGTAAACAAGTTGTGAGGAGTGGACAGACAGTGGAGGCTTGGG 712  
Db 630 ACAATCTCTCTCAGGTAAACAAGTTGTGAGGAGTGGACAGACAGTGGAGGCTTGGG 689  
QY 713 AAGAAACATCGATAGGATGTTCAAGGATGCAATCACTCTGGAAGCGTGTCCGAGCGA 772  
Db 690 AAGAAACATCGATAGGATGTTCAAGGATGCAATCACTCTGGAAGCGTGTCCGAGCGA 749  
QY 773 GACCAACATCATCTGCTGCTGTTTCTGTCAGTTTCAAGCGCGGTGTAACCTGGCGCTGAT 832  
Db 750 GACCAACATCATCTGCTGCTGTTTCTGTCAGTTTCAAGCGCGGTGTAACCTGGCGCTGAT 809  
QY 833 GAAGCTCCGAGGTCTGAGACACTGTGCTAGCTAGGATGCTTCTGATTTGCTGCTGTAAT 892



Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Hordeum.  
Bougrì.O. and Grimm.B.  
Members of a low-copy number gene family encoding glutamyl-tRNA  
reductase are differentially expressed in barley  
Plant J. 9 (6), 867-878 (1996)  
96267576  
PUBMED  
REFERENCE  
2 (bases 1 to 1997)  
Bougrì.O.  
Direct Submission  
Submitted (07-APR-1995) O. Bougrì, Institute of Plant Genetics &  
Crop, Plant Research, Corrensstr. 3, 06466 Gatersleben, FRG  
revised by author 23-Oct-95  
Location/Qualifiers  
1..1997  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Klages"  
/db\_xref="taxon:112509"  
/chromosome="5"  
/clone="bha87"  
/clone\_lib="lambda gt11"  
/dev\_stage="12 days old seedlings"  
1..1997  
/gene="hema 1"  
1..1997  
/gene="hema 1"  
/evidence=experimental  
333..1919  
/gene="hema 1"  
/function="aminolevulinic acid synthetase (chlorophyll  
synthesis)"  
/note="1st isoform"  
/codon\_start=1  
/product="glutamyl tRNA reductase"  
/protein\_id="CAA60054.1"  
/db\_xref="GI:1039332"  
/db\_xref="SWISS-PROT:Q42843"  
/translation="MMAGATSAATAAGAAKARGPAACAAKRGVRRSSGVVRVC  
DAGDAQAASKAASITALEQPKISADRYMKKKSIYAVIGLSVHTAPVMREKLVAABE  
LWPRALSELTSLNHLEEARVLSTCNMEIYYVALSWNRGIRVDVDMSKSGIPASEL  
REHLFWLRSDATRHILFVEAGSLDLVIGEGQILAQQVVNRNGSQSGIGKNIDRMFL  
KDAILTAGKRVRCETNITSAGAVSVSSAAVELAMMLPKSECLSRMILLTGAGQGKLV  
KHLIAAGCKKVVPVNSVERVDIAIEEMKDIEIVYRPLETEMYEAADAADVFTSTASE  
SLFFTEHAELVPPISLWAGVGRLPVDSIPRNVGACLSEVEHARVYNDDLKVEVVEA  
NKEDVRKAMEAQTIITOLKRFRAWRDSLTPVTIKLRSVADIRASELEKCLKIKI  
GEDNLNKKMRSSIELSGLIYVNKLHGPIQLHLCRDGSDSRITDETLENHNLNRMFSL  
DTEKAVLEQKIYKAKVKETQS"  
transit\_peptide 333..464  
/gene="hema 1"  
/evidence=experimental  
465..1916  
/gene="hema 1"  
/product="unnamed"  
/evidence=experimental

BASE COUNT 504 a 503 c 579 g 411 t  
ORIGIN

Query Match 63.2%; Score 1215.6; DB 8; Length 1997;  
Best Local Similarity 84.9%; Pred. No. 6.3e-199;  
Matches 1373; Conservative 0; Mismatches 235; Indels 9; Gaps 1;

QY 70 TTAATGGCAACCAGCGTCAGCGAACCCGCCGCAGCAGCCGCCACCCACCGCCAAGC 129  
Db 331 TGATGATGGCGGAGCGACGTGACGACCGCGAGCGCGCGCATTCGCCCGCCGAAGG 390  
QY 130 CGCGGGGGTCTGTCGTGGCGCCCTCTGCCAGAGGGTGGCCGGCGCGGAGCGCGCTCCG 189  
Db 391 CGCGGGGGCCCGCGCACAGCTGCCCTCGCTCGTCCGCGCGGGCGCGGAGGGGTCCG 450

	Matches 1373;	Conservative	0;	Mismatches 235;	Indels	9;	Gaps	1;
QY	70	TAATGGGACACGAGCTCAGGACACCGCGCGCGAGCACCACCGCGAAGC						129
Db	331	TCATATGGCGGACGACGTACGCCAGGACCGCGCGGCATTCGCGCGCGCAAGG						390
QY	130	CGCGGGGTGCTGTCGCGCCTCTGCCAGAGGGTGGCGCGCGGCGGCGGCGCTCCG						189
Db	391	CGCGGGGCGCGCGCACGCTGCCCTCGCTCGCGCGCGCGCGCGGCGAGCGGTCCG						450





REFERENCE AUTHORS Ehrhartidae; Oryzae; Oryza.  
TITLE Polynucleotides encoding aminolevulinic acid biosynthetic enzymes  
JOURNAL Patent: WO 0109304-A 7 08-FEB-2001;  
E.I. DU PONT DE NEMOURS AND COMPANY (US)  
FEATURES Location/Qualifiers  
source 1..1778  
/organism="Oryza sativa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4530"  
BASE COUNT 496 a 386 c 479 g 417 t  
ORIGIN  
Query Match 62.2%; Score 1196.4; DB 6; Length 1778;  
Best Local Similarity 87.0%; Pred. No. 1.3e-195;  
Matches 1314; Conservative 0; Mismatches 196; Indels 0; Gaps 0;  
QY 210 GCCGCGTGGAGGCCAGCGCCGCTGCGCAAGGCTGCCAGGCTCGCCGCTCGAG 269  
DB 1 GCACAGTGGTACCCAGCGCGAGCGGTGGCCAAAGCGCGCTCGCCGCTCGAG 60  
QY 270 CAGTTCAAGATATCCGCGACCGGTACATGAAGAAAGAGTACCATAGCTGTGATGCC 329  
DB 61 CAGTTCAAGATCTCCGCGACCGGTACATGAAGAAAGAGTACGATAGCGGTATAGGC 120  
QY 330 CTCAGTGTACACAGCACCGAGTGGAGATGGGTGAAACCTTGTTCACAGAGAACTG 389  
DB 121 CTCAGTGTACACACTGCACCGAGTGGAGATGGGTGAAACCTTGTTCACAGAGAACTA 180  
QY 390 TGGCCCGCTGTATTCAAGAACTCACTAGCTGCAACCATATTGAGAGGCTGCTGTTCT 449  
DB 181 TGGCCCGCTGTATTCTCAAGTACCAGTCTCAATCATATTGAAGAGGCTGCTGTTCT 240  
QY 450 AGTACCTGTATAGAAATTTATGTGGCGCTTATCATGGAACCGTGGTATCAGA 509  
DB 241 AGTACCTGCAATAGAAATCTATGTGTAGCTTTATCGTGAACCGTGGATTAGA 300  
QY 510 GAAGTAGTGGACTGGATCTCGAAGAAAGTGGTATTCGGCTTCGAGCTTAGGAGCAC 569  
DB 301 GAAGTGGTAGACTGGATCTCAAGAAAGAGTGGATTCCTGCTTCTGAGCTCAGGAGCAT 360  
QY 570 CTGTTTCATCTTGGCAAGCAGTGATGCCACAGCCATCTGTTGAGGTGTCAGCTGGCCT 629  
DB 361 CTATTCATGTTGCGTGACAGTATGTCACAGCCATCTGTTGAGGTATCTGCTGGGCT 420  
QY 630 GACTCTTTGTTCTCGGTGAAGGACAAATCTTGTCTCAGGTAAACAAAGTTGTGAGGAGT 689  
DB 421 GACTCTTTGTTCTTGGAGAGGCGCAATCTTGTCTCAAGTTAAACAAAGTTGTGAGAGT 480  
QY 690 GGACAGACAGTGGAGGCTTGGGAAGAACATCGATAGATGTTCAAGGATGCAATCACT 749  
DB 481 GGGCAAAACAGTGGAGGCTTGGGAAGAACATCGATAGATGTTCAAGGATGCAATCACT 540  
QY 750 GCTGGAAGCGTGTCCGCGAGCAGACCAACATATCATCTGCTGCTGTTTCGTCAGTTCA 809  
DB 541 GCTGGAAGCGTGTCCGCTGCGAGACTAACATATCATCAGTGTCTGCTCTGTCAGTTCA 600  
QY 810 CGCGGGTGTAACTGGCCCTCATGAGCTCCGAAAGTCTGAAGCACTGTCTAGCTAGGATG 869  
DB 601 GCTGAGTGTAAATGGCCCTTGTATGAAGCTTCCAAAGTCCGAATGCTATCTGTAGGATG 660  
QY 870 CTCTGATTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 929  
DB 661 CTGTTGATTGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 720  
QY 930 TGCAGAGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 989  
DB 721 TGCAGAGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 780  
QY 990 ATGAAGATATAGAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1049

DB 781 ATGAAGACATTTGAGATTGTTGATCAGGCTCTTACAGAGATGTATGAAGCCGCTGCCGA 840  
QY 1050 GCTGATGTGTTTACAGACACCGCATCTGAAACTTCAATTTGTCGAAAAAACAACGCA 1109  
DB 841 GCTGATGTGTTTACAAAGCAGCGCATCCGAAACCCCATTTGTTTACAAAGAGGACGCA 900  
QY 1110 GAGGCACTCCCGCTGTCTCTGATCTATGAGAGGTGTTGCGCTGTTGTTGTCGATATCT 1169  
DB 901 GAGGCGCTTCCCGCTATTTCTGATGCTATGGTGGTGTTCGACTCTTTGTGACATATCC 960  
QY 1170 GTCCCCAGGAATGTACGCGCATGTGTCTGAAGTTGGCGCTGACACGAGTGTACAAATGTC 1229  
DB 961 GTCCCCAGGAATGTACGCGCTGTGTCTGAAGTTGGCCATGCGCGAGTATACAACTGC 1020  
QY 1230 GACGACTGAAGAGGTGGTGAAGCAACAAAGAGGACCGGCTCAGAAAGCAATGAG 1289  
DB 1021 GATGACTTGAAGAGGTGTGGAAGCCAAACAAAGGAGGACCGCTTAGGAAAGCAATGAG 1080  
QY 1290 GCGCAGACATCATCACCGAAGACTGAGAGGTTTCGAGGATCGAGGAGTCTCGCTGAG 1349  
DB 1081 GCCCAACAAATCATCACCCCAAGAAATTGAACCGTTTCGAGGATCGAGGAGTCTCGCTGAG 1140  
QY 1350 ACCGTTCCGACCATCAAGAACTGAGGTCTGACGCGACAGATCAGGCGCTTCGAGGCTC 1409  
DB 1141 ACTGTTCCGACTATCAAGAGCTGAGGTCTTACGCGACAGATCAGGCGCTTCGAGGCTT 1200  
QY 1410 GAGAAGTGCCTGCAAGAAAGTAGGTGAGGACCGCTTCAACAAAGAGATGAGGAGGCCATC 1469  
DB 1201 GAGAAGTGCCTTCCAGAAAGTCCGCGAAGAGCGCCCTCACCAGAAAGATGAGAAGATCCATC 1260  
QY 1470 GAGGAGTGGACACCGGCTGTTAAACAGTCTCTCATGTCGCGCTGCGAGCCTGAGG 1529  
DB 1261 GAGGAGTGGACACCGGCTGTTAAACAGTCTCTCATGTCGCGCTGCGAGCCTGAGG 1320  
QY 1530 TGGCAGCGCAGCAGCAGCGCCCTTGAACGAGAGCTCGAGAAACATCAGCGCCCTCAAC 1589  
DB 1321 TGTGACGCGCAGCAGCAGCGCCCTCGATGAGAGCTGAGAGAACATGCAAGCTTCAAC 1380  
QY 1590 CGGATGTTGAGCTTCGACATGAGAGGCGGATCATCGAGCAGAGATCAAGGCCAAGGTG 1649  
DB 1381 AGGATGTTGAGCTTCGACACCGAGAGGCGATCATTTGAGCAGAGATCAAGGCCAAGGTG 1440  
QY 1650 GAGAGACACAAACTGAGGCGGAGGAGCAATTTTTCACCACTATCTATATATA 1709  
DB 1441 GAGAAGTCCCAAGAACTGAGATTGAAGAGAGATTTTTTTTTTTCAGCCCGCTGATCTACT 1500  
QY 1710 GCGTCTCCAA 1719  
DB 1501 ATGTATACTA 1510  
RESULT 7  
AX081089  
LOCUS Sequence 19 from Patent WO0109304.  
DEFINITION AX081089  
ACCESSION AX081089  
VERSION AX081089.1 GI:13170008  
KEYWORDS  
SOURCE  
ORGANISM  
Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Triticum.  
REFERENCE 1  
AUTHORS Cahoon,R.E., Gutteridge,S., Harvell,L.T., Rafalski,J.A., Tao,Y. and Weng,Z.  
TITLE Polynucleotides encoding aminolevulinic acid biosynthetic enzymes  
JOURNAL Patent: WO 0109304-A 19 08-FEB-2001;  
E.I. DU PONT DE NEMOURS AND COMPANY (US)  
FEATURES Location/Qualifiers  
source 1..1480  
/organism="Triticum aestivum"  
/mol\_type="genomic DNA"

Db	966	AGGAAGACCGTGTGAGGAAGCAATGAGGCCCAACAATCATTTACCCGAAGTGAAC	1025
Qy	1321	GGTTCGAGGCGATCGAGGGACTCGCTCGAGACCGTTTCGACCATCAAGAAGCTGAGTCTGT	1380
Db	1026	GGTTCGAGGCGATCGAGGGACTCACTCGAGACCGTTTCGACCATCAAAAAGCTGAGTCTGT	1085
Qy	1381	ACGCGGACAGGATCAGGGCCCTCGAGCTCGAGAAGTGCCTGCAGAAAGTAGGTGAGGACG	1440
Db	1086	ACGCGGACAGGATCAGGGCCATCGAGCTCGAGAAGTGTCTGCAGAAAGTTCGGGAAGACA	1145
Qy	1441	CCCTCAACAAAGATGAGGAGAGCCATCGAGGAGCTGAGCACCGGCATCGTTTAAACAAGC	1500
Db	1146	ATCTCAACAAAGATGAGAGGTCCATCGAGGAGCTGAGGACCGGCATAGTGAAACAAGC	1205
Qy	1501	TCCTCCATGCGCCGCTCGAGCAGCTGAGTGCAGCGGAGGACGACCGCCACCTTGAGC	1560
Db	1206	TCCTTCACGCGCCCACTGCAGCACCTGAGATGCGACGCGGACGACCGCCACCTTGAGC	1265
Qy	1561	AGACGCTCGAAGACATGACGCGCCCTCAAACCGGATGTTTCAGCCCTCGACATGGAAGAGCGCA	1620
Db	1266	AAAGCTTGAGACATGACGCGCTCAACAGATGTTCAACCTCGACACGGAGAGCGCG	1325
Qy	1621	TCATCGAGCAGAAGATCAAGGCCAAGGTGAGGAGACACAAATCTGAGGCCAGGAAGCAA	1680
Db	1326	TCCTTGAGCAGAAGATCAAGGCCAAGGTGAGGAGACACCAAGCTGAGACCGAGGACAC	1385
Qy	1681	TTTTTTCACACCAATTCATCTATATATAGGCTCTCCAATCTCATTCATTTT	1733
Db	1386	TTGCGCGTC--TGTATATCTACTTATCTACTGCTCCAGAGATGCTGCTACATTTCT	1436
RESULT 8			
D88383			
LOCUS	D88383	Hordeum vulgare hema3 mRNA	1822 bp mRNA linear PLN 17-MAR-1998
DEFINITION		cds.	
ACCESSION	D88383		
VERSION	D88383.1	GI:2967442	
KEYWORDS	hema3; glutamyl-tRNA reductase.		
SOURCE	Hordeum vulgare subsp. vulgare		
ORGANISM	Hordeum vulgare subsp. vulgare		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
	Pooideae; Triticeae; Hordeum.		
REFERENCE	1 (bases)		
AUTHORS	Tanaka, R., Yoshida, K., Nakayashiki, T., Tsuji, H., Inokuchi, H., Okada, K. and Tanaka, A.		
TITLE	The third member of the hema gene family encoding glutamyl-tRNA reductase is primarily expressed in roots in Hordeum vulgare		
JOURNAL	Photosyn. Res. 53, 161-171 (1997)		
REFERENCE	2 (bases 1 to 1822)		
AUTHORS	Tanaka, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-OCT-1996) Ryouichi Tanaka, Kyoto University, Department of Botany, Faculty of Science, Sakyo-ku, Kitashirakawa-oiwakecho, Kyoto, Kyoto 606-01, Japan		
	(E-mail: ryouichi@bot.kyoto-u.ac.jp, Tel: 075-753-4147, Fax: 075-753-4122)		
FEATURES			
source	Location/Qualifiers		
	1..1822		
	/organism="Hordeum vulgare subsp. vulgare"		
	/mol_type="mRNA"		
	/sub_species="bonus"		
	/db_xref="taxon:112509"		
	/clone="BHA200"		
gene	1..1822		
	/gene="hema3"		
CDS	28..1635		
	/gene="hema3"		
	/note="isoform 3"		
	/codon_start=1		
	/product="glutamyl-tRNA reductase"		

Query Match	57.2%	Score 1100.4	DB 8	Length 1822
Best Local Similarity	78.9%	Pred. No. 3.9e-179		
Matches 1311	Conservative 0	Mismatches 351	Indels 0	Gaps 0
Qy	76	CGACACGAGCGTCAAGCAGCACACCGCGCGGCGAGCAGCGCGCCACACCGCCAAAGCGCGGG	135	
Db	38	CGAGCAGGCTCCGCGAGCGCCATGCGCGGCGGTTGCGCGCGCGCGGTGCACAAAGC	97	
Qy	136	GGTGCTGTCGGCCCTTCGCGAGAGGTGGCGCGCGCGGCGAGCGCGCTCCGGGGTGG	195	
Db	98	CCGAGAGCTTCGCTGCTTCGCGAGGGTGGCGCGCGCGCGCGCGCGCGCTGTCTCTGCG	157	
Qy	196	TGCGGTGCGAGCGCGCGCGGTGAGCGCCAGCGCGCAGCGCGTGGCGAAAGCTGCCAGCG	255	
Db	158	TCGTGCGGTGGACGCGCGCGCGCGGGGTCCGGCCAGATGGCGGCCCATGGCTGCCAGCG	217	
Qy	256	TCGCGCGCTCGAGCAGTTCAAGATATCCGCGCAGCCGGTACATGAAGAAAGAGATACCA	315	
Db	218	TCGCGCGCTCGAGCAGTTCAAGATCTCCGCGCAGCTACATCAAAAGAAAGAGTAGCA	277	
Qy	316	TAGTGTGATAGGCTCAGTGTACACACAGCACAGTGGAGATGCGTGAAGAACTTTGCTG	375	
Db	278	TTGCTGTCTAGGACTCAGTATCCACACAGCACCGGTGAGATGCGTGAAGAAAGCTTGGCG	337	
Qy	376	TTGCAGAGGAACTGTGGCCCCGTCTATTCAAGAACTCAGTACCTGCAACCATATTGAAG	435	
Db	338	TCGCAGAAAGACTATGCCCCCGTGTTCGAGAGCTAACCAATCTTAATCATATTGAAA	397	
Qy	436	AGGCTGCTGTTCTTAGTACCTGTAATAGAATGAAATTTATGTGGTGGCGCTATCATGGA	495	
Db	398	AGGCTGCTGTTCTTAGCCCCCTGCAATCGGATGAAATTTATGTGGTGGCTTTATCCTGGA	457	
Qy	496	ACGCTGATACAGAGAGTAGTGGACTGGATGTCGAAGAAAGTGGTATTCGCCGCTTCG	555	
Db	458	ACGCTGGCATTAGAGAGATAGTGATTTGATGTCATGAAGAAAGCGGATCCCTGCTGTTG	517	
Qy	556	AGCTTAGGAGACACTCTTTCATCTTCGGAAGCAGTGTATGCGCACACCGCATCTGTTTGAGG	615	
Db	518	AGCTCAGGAGACCTCTTTCATGTTCCGTGACAGTGTGCGCACACGCCATCTGTTTGAGG	577	
Qy	616	TGTCAGTGGCTTGACTCTTTGGTCTCCGTTGAAGACAAATTCCTTGCTCAGGTTAAAC	675	
Db	578	TTTCTTCGGGCTCGACTCTTTGGTCTTCGAGAAAGACAGATCTTGCTCAGGTTAAAC	637	
Qy	676	AAGTTGTGAGAGTGGACAGAACAGTGGAGGCTTGGGAAGACATCGATAGATGTTCA	735	
Db	638	AGGTTGTCAAGTGGCAAAAAGTGGAGGCTTGGGAAGACATTTGATAGGATGTTCA	697	
Qy	736	AGGATGCAATCACTGTGGAAGCGGTGTCGCGAGCGAGACCAACATATCATCTGTTGCTG	795	
Db	698	AGGATGCAATCAGCAGGAAAGCGGTCGCTCTGAGACCAACATATCATGTGGCGCG	757	
Qy	796	TTTCTGTGACTCAGCGGCGTTGAACTGCCCTGTATGAAGCTTCGAGTGTGAAGCAC	855	
Db	758	TTTCTGTGACTCAGCTGCGGTGAAGATTAAGCCCTGTATGAAGCTTCCAAAGCTCTGAAATGCT	817	
Qy	856	TGTCAGCTAGGATGCTTCTGATTTGGTCTGGTAAAATGGAAAGCTAGTGATCAAAATC	915	

**TITLE**  
 Members of a low-copy number gene family encoding glutamyl-tRNA  
 reductase are differentially expressed in barley  
**JURNAL**  
 Plant J. 9 (6), 867-878 (1996)  
**REFERENCE**  
 MEDLINE  
 96267576  
 8696365  
**AUTHORS**  
 Bougri, O. and Grimm, B.  
**TITLE**  
 Members of a low-copy number gene family encoding glutamyl-tRNA  
 reductase are differentially expressed in barley  
**JOURNAL**  
 Unpublished  
**REFERENCE**  
 3 (bases 1 to 1703)  
**AUTHORS**  
 Bougri, O.  
**TITLE**  
 Direct Submission  
**JOURNAL**  
 Submitted (07-APR-1995) O. Bougri, Institute of Plant Genetics &  
 Crop, Plant Research, Corrensestr. 3, 06466 Gatersleben, FRG  
**FEATURES**  
 Location/Qualifiers  
 1..1703  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cui="taxon:112509"  
 /db\_xref="taxon:112509"  
 /chromosome="5"  
 /clone\_lib="lambda ZAPII"  
 /dev\_stage="6 days old etiolated seedlings"  
 1..1703  
 /gene="hemA 2"  
 <1..1399  
 /gene="hemA 2"  
 /function="aminolevulinate synthesis (chlorophyll  
 synthesis)"  
 /note="2nd isoform"  
 /codon\_start=2  
 /product="glutamyl tRNA reductase"  
 /protein\_id="CAA0055.1"  
 /db\_xref="GI:1666079"  
 /db\_xref="SWISS-PROT:Q96563"  
 /translation="QFISADRYIKKESIAVIGLSVHTAPVDMREKLAVALMPRA  
 ISLSTLSNHIIEAAVLSTCNRMETIYVVALSWNRIGREVVDWMSKSGIPASELREHLF  
 MLRDSGATRIHFEVSAGLSLVLGEGOILAQKVQVWPNGQSGGLKRNIDRMFKDAIT  
 AGKRCATRIISAGVSVSSAAVELAMKLPKSECLARMLLIAGKMGKLVVKKHLIA  
 KGCKVVVNRVVERVDAREMKDIEIVRPLUTEYEAADADVDVFTSASLSLFT  
 KEHAALPPIISLAWGGVRLFDVSPRNVGACLSQVBEHARVYVNDLKEVVEANKEDR  
 VRKAMEAQAIITQELKRFEARWDSLETPTIKLRSYADRIRASELDKCLKQIGENL  
 NKTRRSIEBELSTGIVNKLHPLQHLRCDSRSLDETLDNHNHARNFNDTEKA  
 VLEQIKAKVEKQS"  
 BASE COUNT 466 a 393 c 463 g 381 t  
 ORIGIN  
 Query Match 56.7%; Score 1090.6; DB 8; Length 1703;  
 Best Local Similarity 84.5%; Pred. No. 1.9e-177;  
 Matches 1225; Conservative 0; Mismatches 224; Indels 0; Gaps 0;  
 QY 269 GCAGTTCAAGATATCCGCCACCGTACATGAGGAAAGGATACCATAGCTGTGATAGG 328  
 DB 1 GCAGTTCAAGATATCCGCCACCGTACATCAAGGAAAGGATACCATAGCTGTGATAGG 60  
 QY 329 CCTCAGTGTACACAGCACACAGTGGAGATGCGTGAAGAACTTCTGCTTCACAGGAACT 398  
 DB 61 CCTCAGTGTACACAGCACACAGTGGAGATGCGTGAAGAACTTCTGCTTCACAGGAACT 120  
 QY 389 GTGGCCCGGTGCTATTCAAGAACTCACTAGCCCTGAACCATATTGAAGAGGCTGCTGTCT 448  
 DB 121 ATGGCCCGGTGCTATTTCAGAACTCACCACTGTCGTAACCATCGAAGAGGCTGCTGTCT 180  
 QY 449 TAGTACTGCTAAAGATGAATTTATGTGGTGGCGCTATCATGGAAACCGTGGTATCAG 508  
 DB 181 TAGTACTGCTAAAGATGAATTTATGTGGTGGCGCTATCATGGAAACCGTGGTATCAG 240  
 QY 509 AGAAGTAGTGCATCGGATGTCGAAGAAAGTGTATTCCCGCTTCCGAGCTTAGGGAGCA 568  
 DB 241 AGAAGTAGTGCATCGGATGTCGAAGAAAGTGTATTCCCGCTTCCGAGCTTAGGGAGCA 300  
 QY 569 CTTGTTCACTTCGCGAAGCAGTGCATGCCACGCCCATCTGTTTGAAGTGTCACTGGCGCT 628

DB 301 TCCTTTATGTTGGTCAAGTGTGTCACACGCCATCTGTTGAGGTATCGCTGGGCT 360  
 QY 629 TGACTCTTTGGTCTCGGTGAAGGACAAATCCTTGTCTCAGGTTAAACAAGTTGTGAGAG 688  
 DB 361 CGACTCTTTGGTCTTCTGGGAAGGACAAATCCTTGTCTCAGGTTAAACAAGTTGTAGAAA 420  
 QY 689 TGGACAAACAGTGGAGGCTTGGGAAAGAACATCATAGGATGTTCAAGGATGCATCAC 748  
 DB 421 TGGGCAAAATAGTGGAGGCTTGGGAAAGAACATTTGATAGGATGTTCAAGGATGCAATCAC 480  
 QY 749 TGTGGAAGAGCTGTCGCGACGAGACCAACATATCATCTGGTCTGTTCTGTCTCAGTTTC 808  
 DB 481 GGCTGGAAGAGCGCGCCGCTGCGAGACCAACATATCATAGTGTGTCGCGGTCTGTCTCAGTTTC 540  
 QY 809 AGCGCGGTTGAACCTGCGCTGATGAAGCTTCCGAAGTCTGGAAGCACTGTCTCAGCTAGGAT 868  
 DB 541 AGCGCGAGTTGAGTTGGCCATGATGAAGCTTCCAAAGTCCGAATGCTTGTCTCAGCGAGAT 600  
 QY 869 GCTTCTGATTTGGTGTGTTAAATGGGAAAGCTAGTATCAACATCTGTTGTGCAAAAGG 928  
 DB 601 GCTTTTGAATTTGGCGCGCGCAAAATGGGGAATTTAGTGTCTCAACATTTGATTTGCCAAAGG 660  
 QY 929 ATGCAAGAAAGTGTGTTGTGTGAACCGCTCCGTGGAAGAGGTGATGCTATTCTGTGAGGA 988  
 DB 661 ATGCAAGAAAGTGTGTTGTGTGAACCGCTCCGTGGAAGAGGTGATGCTATTCTGTGAGGA 720  
 QY 989 GATGAAGATATAGAGATCGTGTACAGGCTCTCTCTCAGACATGTTATCAAGCTGTCTGCTGA 1048  
 DB 721 GATGAAGATATAGAGATCGTGTACAGGCTCTCTCTCAGAGATGTTATGAAGCCGCTGCTGA 780  
 QY 1049 AGCTGATGTCGTGTTTACAGACCGCATCTGAAACTTCTATTGTTGGCAAAAGAACACGC 1108  
 DB 781 TGCGCAGCTTGTGTTTACAGACCGCATCTGAACTCTTATTTACCAAGGAGCATGTC 840  
 QY 1109 AGAGGCACTCCCGCTGCTCTGATCTATCTATGGGAGGTGTTCCGCTGTTTCTGCACATATC 1168  
 DB 841 AGAGGCGCTTCTCTCTATCTCTTGGCATGGGTGGTGTGCGGCTTTTCTGCACATATC 900  
 QY 1169 TGTCCCAGGAATGTTCAGCGCATGTGTGCTGAAGTTGGGCTGCACAGAGTGTCAATGT 1228  
 DB 901 CGTCCCAGGAACGTGCGGTGCTGCTATCTCAAGTGGAGCAGCACCGGTATACATGT 960  
 QY 1229 CGAGCACTTGAAGAGGTGTGGAAGCAACAGAGGAGACCGGTCTAGGAAAGCAATGGA 1288  
 DB 961 TGACGACTTGAAGAGGTGTGGAAGCAACAGAGGAGACCGGTCTAGGAAAGCAATGGA 1020  
 QY 1289 GCGCAGACAAATCATCAACGAAGAACTGACACGCTTCGAGGCATGGAAGGAGCTCGCTGGA 1348  
 DB 1021 GGCCCAAGCAATCATCAACCAAGAACTGAAAGCGGTTGAGGCGTGGAGGAGCTCGCTGGA 1080  
 QY 1349 GACCGTTCCGACCATCAAGAACTGAGGTGCTGACGCGACAGGATCAGGCGCTCGAGCT 1408  
 DB 1081 GACCGTTCCGACCATCAAGAACTGAGGTGCTGACGCGACAGGATCAGAGCATCCGAGCT 1140  
 QY 1409 CGAAGAGTCTCGAAGAGTGTGAGGACGCGCTCTCAACGAAGATGAGGAGGAGCAT 1468  
 DB 1141 CGAAGAGTCTCGAAGAGTGTGAGGACGCGCTCTCAACGAAGATGAGGAGGAGCTCTAT 1200  
 QY 1469 CGAGGAGTGTGAGCAGCGGATCGTTAAAGCTCTCTCAAGCGCTGCGAGGAGCTGAG 1528  
 DB 1201 CGAAGAGTGTGAGCAGCGGATCGTTAAAGCTCTCTCAAGCGCTGCGAGGAGCTGAG 1260  
 QY 1529 GTGCGACGCGCAGCAGCGCATCCCTTGAACGAGCGCTCGAAGCAATGCACGCGCTCAA 1588  
 DB 1261 ATGCGATGCGCAGCAGCAGCGCATCCCTTGAACGAGCGCTCGAAGCAATGCACGCGCTCAA 1320  
 QY 1589 CCGGATGTTGAGCTCGAATGGAAGAGGAGTATCTGAGCAGAAAGATCAAGGCCAAGGT 1648  
 DB 1321 CAGGATGTTCAACCTTTGACACGAGGAGGCGGTCTCTGAGCAGAAAGATCAAGGCCAAGGT 1380  
 QY 1649 CGAAGAGACACAAACTGAGGCGCAGGAGCAATTTTCTTCTACCACCATTTCTATATATAT 1708



Db 1381 AGAGAAGCAACCAAGCTGAGGACGAGGACGCTCCGCTTCTGTATATCTACTTATAC 1440  
 Qy 1709 AGCGTCTCC 1717  
 Db 1441 CACCCACACC 1449

# RESULT 10 AC078840/c

LOCUS AC078840 130058 bp DNA linear PLN 07-SEP-2002  
 DEFINITION Oryza sativa chromosome 10 BAC OSJNBb0073N24 genomic sequence,  
 complete sequence.

ACCESSION AC078840

VERSION AC078840.6 GI:22758306

KEYWORDS HTG.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.

## REFERENCE AUTHORS

Buell, C.R., Yuan, Q., Moffat, K.S., Hill, J.N., Jenkins, C.N.,  
 Burr, P.C., Hsiao, J., Zismann, V., Pai, G., Bowman, C.L., Fujii, C.Y.,  
 VanAken, S.E., Bowman, C.L., Craven, B., Utterback, T.R., Khalak, H.,  
 Feldblyum, T.V., Quackenbush, J., White, O., Salzberg, S.L. and  
 Fraser, C.M.

TITLE Oryza sativa chromosome 10 BAC OSJNBb0073N24 genomic sequence

## JOURNAL REFERENCE

2 (bases 1 to 130058)

Unpublished

Buell, R.

## AUTHORS TITLE

Direct Submission

Submitted (07-AUG-2000) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

## JOURNAL REFERENCE

3 (bases 1 to 130058)

Buell, R.

## AUTHORS TITLE

Direct Submission

Submitted (22-SEP-2000) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA, rbuella@igr.org

## JOURNAL REFERENCE

5 (bases 1 to 130058)

Buell, R.

## AUTHORS TITLE

Direct Submission

Submitted (07-SEP-2002) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA, rbuella@igr.org

## AUTHORS TITLE

On Sep 7, 2002 this sequence version replaced gi:10140779.

## COMMENT

Address all correspondence to: rice@tigr.org  
 BAC clone OSJNBb0073N24 is from Oryza sativa chromosome 10  
 The orientation of the sequence is from SP6 to T7 end of the BAC  
 clones.  
 Genes were identified by a combination of several methods: Gene  
 prediction programs including Genscan and Genscan+ (Chris Burge,  
<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkHMM (Mark Borodovsky,  
<http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer  
 (Michael Pertea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)),  
 searches of the complete sequence against a peptide database and  
 the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>).  
 Annotated genes are named to indicate the level of evidence for  
 their annotation. Genes with similarity to other proteins are named  
 after the database hits. Genes without significant peptide  
 similarity but with EST similarity are named as unknown proteins.  
 Genes without protein or EST similarity, that are predicted by more  
 than two gene prediction programs over most of their length are  
 annotated as hypothetical proteins. Genes encoding tRNAs are

predicted by tRNAscan-SE (Sean Eddy,  
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are  
 identified by RepeatMasker (Arian Smit,  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

This BAC overlaps with rice BAC OSJNBa0078001 (AC079888) and  
 OSJNBa0062C05 (AC084023).

## FEATURES source

Location/Qualifiers  
 1. .130058  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="10"  
 /map="S21174S"  
 /clone="OSJNBb0073N24"  
 /notes="japonica cultivar-group"  
 complement(311..1125)  
 /gene="OSJNBb0073N24.23"  
 /note="predicted by Genscan"  
 complement(join(311..452,944..>1125))  
 /gene="OSJNBb0073N24.23"  
 complement(join(311..452,944..1125))  
 /gene="OSJNBb0073N24.23"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAGI3622.1"  
 /db\_xref="GI:10140792"  
 /translations="MVMVTAALLLTDMAGVSYGRRLIPDLDAVAVVGGSPPAKGGVM  
 SRLQVPSPDSGHHVGDQNPYSQNFKIVDPKVFMAAQTLQLQSLTRTTRISSKCNLSLS  
 QVGGK"  
 complement(3145..3192)  
 /rpt\_family="AT\_rich"  
 complement(3862..6664)  
 /gene="OSJNBb0073N24.1"  
 /notes="nearly identical to glutaryl tRNA reductase  
 GI:2920320 (Oryza sativa) (Plant Physiol 117, 332  
 (1998)); EST AU056848,D46763, AU056849, AU056803 from this  
 gene"  
 complement(join(3862..5094,5230..5471,6430..>6664))  
 /gene="OSJNBb0073N24.1"  
 complement(join(3961..5094,5230..5471,6430..6664))  
 /gene="OSJNBb0073N24.1"  
 /codon\_start=1  
 /product="putative glutamyl-tRNA reductase"  
 /protein\_id="AAGI3620.1"  
 /db\_xref="GI:10140790"  
 /translations="MASTTSATAGGAFAAKTRAGSSAAGGACARVAAGRRRSVG  
 VPCDAGVEAQAOQAVAKASVAALQPKISADRYNKERSSTIAVIGLSVHTAPVEMR  
 EKLAVAEELMPTAISLTSLNHIEAAVLTNCNMEIYVVALSNRGIRVVDMMKK  
 SGIPASSELREHFLMRSDATRHLEFVSAGLSVLGEGQILAQVKQVRSQNSGGL  
 GKNIDRMFKDAITAGKRVCCETNISSGAVSVSSAAVELMLKLPKSECLSAARMLLIGA  
 GKMKLVVXHLIAKCKKVVVNRVSDAIREEMKDIIVYRPLTMYEAAAEADV  
 VFSTASEPLFTKEHAELPAISDAMGGVRLFDLSVPRNSACVSEVGHARYVVD  
 DLKVVTEANKEDRLKAMEAQTIITQELKRFPEARWDSLETVPYTIKLRSAADRIARASE  
 LEKQKIGEDALTQKRRSIIETSLGVNKLHLGHLQHLRCDGSDSRITDLETLENMH  
 ALNRMFLDTEKAIIEQTKAKVEKSN"  
 complement(6480..6654)  
 /rpt\_family="(CGG)n"  
 complement(7437..7586)  
 /rpt\_family="Galgin\_Ol2 MITE element from gb:U72728 Oryza  
 longistaminata receptor-like kinase protein (Xa21), family  
 member F, pseudogene sequence (233 to 384) 152 nt"  
 complement(7443..7584)  
 /rpt\_family="Galgin\_Os3 element from gb:D32165 Rice gene  
 for aspartic protease (302 to 448) 147 nt"  
 complement(8050..8133)  
 /rpt\_family="AT\_rich"  
 complement(8285..8320)  
 /rpt\_family="AT\_rich"  
 8873..8995  
 /rpt\_family="Galgin\_Os3 element from gb:D32165 Rice gene  
 for aspartic protease (302 to 448) 147 nt"



repeat_region	8873. .8994 /rpt_family="Gaigin_O12_MITE_element_from_gb.U72728_Oryza_longistaminata_receptor-like_kinase_protein_(xat1),_family_member_F_pseudogene_sequence_(233_to_384)_152_nt" complement(11172. .13291) /gene="OSJNBb0073N24.2" /note="similar to steroid membrane binding protein GI:1657409 (Sus scrofa); EST C26842,D48715, AA231657 from this gene" complement(join(<11172. .11511,12724. .13006,13085. .>13291)) /gene="OSJNBb0073N24.2" complement(join(11172. .11511,12724. .13006,13085. .13259)) /gene="OSJNBb0073N24.2" /codon_start=1 /product="putative steroid membrane binding protein" /protein_id="AAG13623.1" /db_xref="GI:10140793"	Db	5411	 TGGAGATCGTGAGAACTTGTGTTCAGAGAACTATGGCCCCGTCTATCTCAGAAC	5352
gene		Qy	412	 TCACCTAGCCTGAACCATATTGAAGAGGCTGCTGTTCTTTAGTACCTGTATAGATGGA	471
		Db	5351	 TCACCAGTCTGAATCATATTAAGAGGCTGCTGTTCTTTAGTACCTGCAATAGATGANA	5292
mRNA		Qy	472	 TTTATGTGGTGGCGCTATCATCGAAACCGTGGTATCAGAGAACTAGTGGATGATCGA	531
CDS		Db	5291	 TCTATGTGGTGTGCTTTATCGTGGAAACCGTGGGATAGAGAACTGGATGATGATCA	5232
		Qy	532	----- -----	531
		Db	5231	AGTAAAGATTGTTTCATCAAAATGCTCAATTTGTTCTGTCATAGTCTCTTTTCTGCT	5172
		Qy	532	----- -----	531
		Db	5171	TTTCCCTTTTCTGTGACTGAAGAACGGGTTTGTCTTAGATGGTTCTTAATCTCTGCT	5112
		Qy	532	----- -----	576
		Db	5111	 GCTTTTTCCTCAACAGAAAGTGGAAATCCCTGCTTCTGAGCTCAGGGAGCATCTATTCA	5052
repeat_region		Qy	577	 TCTTCGGAAGCAGTATGCCACACGCCATCTGTTTGGAGTGTGCTGCTGCTGCTGCT	636
repeat_region		Db	5051	 TGTTCGCTGACAGTATGCCACACGCCATCTGTTTGGAGTATCTGCTGGCTTGACTCT	4992
gene		Qy	637	 TGGTTCTCGGTGAAGGACAAATCTCTGCTCAGGTTTAAACAAGTTCTGAGAGTGGACA	696
mRNA		Db	4991	 TGGTTCTTGGAGAGGGCAATCTCTGCTCAAGTTTAAACAAGTTCTGAGAGTGGCAAA	4932
CDS		Qy	697	 ACAGTGGAGGCTTGGGAAGAACATCGATAGATGTTTCAAGGATGCTCACTGCTGGA	756
		Db	4931	 ACAGTGGAGGCTTGGGAAGAACATCGATAGATGTTTCAAGGATGCTCACTGCTGGA	4872
		Qy	757	 AGCGTGTCCGAGGAGACCAATATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	816
		Db	4871	 AGCGTGTCCGCTGCGAGACTAACATATCATCAGGTGCTGCTGCTGCTGCTGCTGCTG	4812
		Qy	817	 TTGAACTGGCCCTGATGAAGCTTCGGAAGTCTGAAGCACTGTGCTGCTGCTGCTGCTG	876
		Db	4811	 TTGAACTGGCCCTGATGAAGCTTCGGAAGTCTGAAGTCCGATGCTGCTGCTGCTGCTG	4752
		Qy	877	 TTGCTGCTGTTAAATGGGAAGCTAGTATGATCAACATCTGCTGCTGCTGCTGCTGCTG	936
		Db	4751	 TTGCTGCTGCAAGATGGGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	4692
		Qy	937	 AGGTTGTTGCTGAACCGCTCCGCTGGAAGGGTGGATGCTTATTCGCGAGAGATGAAG	996
		Db	4691	 AAGTTGTTGCTGAACCGCTCCGCTGGAAGGGTGGATGCTTATTCGCGAGAGATGAAG	4632
		Qy	997	 ATATAGAGATCGTGTACAGGCTCTCTCAGACATGTATCAAGCTGCTGCTGCTGCTGCTG	1056
		Db	4631	 ACATTGAGATTGTGTACAGGCTCTTACAGAGATGTATGAGCCGCTGCTGCTGCTGCTG	4572
		Qy	1057	 TCGTGTTACACAGCACCGCTCTGAACTTCACTTGTTCGAAAAAGAAACACGAGAGG	1116
		Db	4571	 TCGTGTTACAAAGCACCGCTATCCGAAACCCCATTTGTTCAAAAGGAGCAGCAGAGG	4512
		Qy	1117	 TCCCCCTGCTCTGATATGAGGAGTGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1176
		Db	4511	 TTCCCGCTATTTCTGATGCTATGAGTGTTCGACTCTTTGTGACATATTCGCTGCTG	4452
		Qy	1177	 GGAATGTACGCGCTGTGTCTGAAAGTGGCGCTGCAACGAGTGTACAAATGCTGCACT	1236
		Db	4451	 GAAATGTACGCGCTGTGTCTGAAAGTGGCGCTGCAACGAGTGTACAAATGCTGCACT	4392
		Qy	1237	 TGAAGAGGTGTGGAAAGCCAAACAGAGAGACCGGCTCAGGAAAGCAATGAGAGCGCA	1296
		Db	4391	 TGAAGAGGTGTGGAAAGCCAAACAGAGAGACCGGCTCAGGAAAGCAATGAGAGCGCA	4332
		Qy	1297	 CAATCATCCGGAAGACTGAGACGTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGGCTT	1356

Query Match . 51.3%; Score 987; DB 8; Length 130058;  
 Best Local Similarity 79.5%; Pred. No. 1.1e-159;  
 Matches 1243; Conservative 0; Mismatches 185; Indels 135; Gaps 1;

Qy 292 GGTACATGAAGAAAGTACCATAGCTGTGATAGGCTCAGTGTACACACACACAG 351  
 |||||  
 Db 5471 GGTACATGAAGAAAGTACGATAGCGGTAAATAGGCTCAGTGTACACACTGCACCG 5412  
 |||||

Qy 352 TGGAGATGCGTGAAGAACTTGTGTTGAGAGAACTGTGCGCCCGTCTATTCAAGAAC 411  
 |||||

Db	4331	CAATCATCACCCAGAAATTGAAA	CGTTTCGAGGCATCGGGAGCTCGCTGGAGACTGTT	4272
Qy	1357	CGACCATCAAGAAGCTGAGTCTG	TACGGCGACAGATCAGGCGCTCGGAGCTCGAGAAGT	1416
Db	4271	CGACTATCAAGAAGCTGAGTCTC	TACGGCGACAGATCAGGCGCTCGGAGCTCGAGAAGT	4212
Qy	1417	GCCTGCAGAAAGTAGTTCAGAGC	CGCTCACCAGAAGATCAGGAGCCATCGAGGAGC	1476
Db	4211	GCCTCCAGAAGATCGCGAAGACG	CCCTCACCAGAAGATCAGGAGCCATCGAGGAGC	4152
Qy	1477	TGAGCACCGGCATCGTTAACAG	CTCTCCATGGCCCGCTGCAGCACCTGAGGTCGACG	1536
Db	4151	TCAGCACCGGCATCGTTAACAG	CTCTCCATGGCCCGCTGCAGCACCTGAGGTCGACG	4092
Qy	1537	CGAGCGACGCGCACCTTGACGAG	CGCTCGAGAACATGACCGCCCTCAACCGGATGT	1596
Db	4091	CGAGCGACGCGCACCTTGACGAG	CGCTCGAGAACATGACCGCCCTCAACCGGATGT	4032
Qy	1597	TCAGCCTCGACATGAGAGGCGAT	CATCGAGCAGAAGATCAAGCCCAAGGTGGAGRAGA	1656
Db	4031	TCAGCCTCGACATGAGAGGCGAT	CATCGAGCAGAAGATCAAGCCCAAGGTGGAGRAGA	3972
Qy	1657	CACAAACTGAGGCGAGGAAGCA	ATTTTCTACCAACCATTTATATATATAGCGTCTC	1716
Db	3971	CCAGAACTGAGATGGAAGAGAG	ATTTTCTACCAACCATTTATATATATAGCGTCTC	3912
Qy	1717	CAA 1719		
Db	3911	CTA 3909		
RESULT 11				
AE017109/c				
LOCUS	AE017109	300029 bp	DNA linear	PLN 06-JUN-2003
DEFINITION	Oryza sativa (japonica cultivar-group) chromosome 10, section 63 of 77 of the complete sequence.			
ACCESSION	AE017109	AE016959		
VERSION	AE017109.1	GI:31432860		
KEYWORDS	Oryza sativa (japonica cultivar-group)			
SOURCE	Oryza sativa (japonica cultivar-group)			
ORGANISM	Oryza sativa (japonica cultivar-group)			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.			
AUTHORS	1 (bases 1 to 300029)			
CONSTRM	The Rice Chromosome 10 Sequencing Consortium			
TITLE	In-depth view of structure, activity, and evolution of rice chromosome 10			
JOURNAL	Science 300, 1566-1569 (2003)			
REFERENCE	2 (bases 1 to 300029)			
AUTHORS	Buell,C.R., Wing,R.A., McCombie,W.R., Messing,J. and Yuan,Q.			
TITLE	Direct Submission			
JOURNAL	Submitted (05-MAY-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
COMMENT	This is the pseudomolecule for rice chromosome 10, which was constructed by resolving discrepancies between overlapping BACs, trimming the overlap regions, and linking the unique sequences to form a contiguous sequence. Genes in individual BAC clone were identified by a combination of several methods: Gene prediction programs, searches of the complete sequence against a peptide database and EST databases. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <a href="http://genome.wustl.edu/eddy/tRNAscan-SE/">http://genome.wustl.edu/eddy/tRNAscan-SE/</a> ).			
FEATURES	Location/Qualifiers			
source	1. 300029			
	/organism="Oryza sativa (japonica cultivar-group)"			
	/mol_type="genomic DNA"			

/cultivar="Nipponbare"	
/db_xref="taxon:39947"	
/chromosome="10"	
<1. 124742	
/note="Chromosome Sequence Derivation: nucleotide sequence in this region was derived from BAC clone OSJNBa0017E08 (GB:AC068923)." "	
complement(<653. .>1075)	
/locus_tag="OSJNBa0017E08.22"	
/locus="predicted by fgenesh"	
complement(<653. .>1075)	
/locus_tag="OSJNBa0017E08.22"	
complement(653. 1075)	
/locus_tag="OSJNBa0017E08.22"	
/codon_start=1	
/product="hypothetical protein"	
/protein_id="AAP54437.1"	
/db_xref="GI:31432861"	
/translation="MSVLSFTFLPSLPSLSSTSPSPSSSTSPFAPAAGASNPPIPARARQVQQVTHGAVGADLHVGAAGHGGLDFVFSRSHAARKLIDLVTLSRARVVTSKQLVSHVKLCILVCRDLDLFLPSAHCCNQAKKVAGQAVAE"	
<2371. .>5902	
/locus_tag="OSJNBa0017E08.8"	
/note="similar to NADPH dehydrogenase GB:CAA59228 GI:683476 (Hordeum vulgare); EST D46920, BF430694, AU096023 from this gene"	
join(2371. .2504,2643. .2729,3928. .4459,5092. .5902)	
/locus_tag="OSJNBa0017E08.8"	
join(2442. .2504,2643. .2729,3928. .4459,5092. .5618)	
/locus_tag="OSJNBa0017E08.8"	
/codon_start=1	
/product="putative dehydrogenase"	
/protein_id="AAP54438.1"	
/db_xref="GI:31432862"	
/translation="MALQAAATTSFLPSALSARKEGAVKDSAFGLVRLDGLKLETSALGLTKRVTSVAIRAQAASVSTPTVPASPSGKOTLRKGTAVITGASGGLGLATAKALAEGRWHVWVWGRDPLKASRAAKAAGMEKGSYTVHLDLSDSVQRPVAVRRL EMPDVVVCAAVYQPTAKQPSFTADGFEMSVGVNHLGHELLARELLADLTSDDYPSK RLIIVSGITGNTLGNVPKANIQLDLGLASGLDGVSSAMIDGGDFGAKAYKOS KVCNMLTMQEPHRRYHGETGTGTFTASLYPGCIATGTLFRHVPLFLPLPPFKYITKG YVSEEGAKRLAQVSDPSLTSGVYWSWNNNSASFENQLSEASDPEKAKKWELSE KLVLGLADHDQ"	
<8942. .>17335	
/locus_tag="OSJNBa0017E08.7"	
/note="similar to dentin phosphorin protein GB:AAD16120 GI:4322670 (Homo sapiens); EST D47284, AA751668, AU088689 from this gene"	
join(<8942. .9423,10749. .10848,11202. .11535,11806. .11918, 12487. .12615,12796. .12889,12992. .13075,13315. .13946, 14712. .15001,15148. .15292,15926. .16023,16125. .16206, 16895. .17335)	
/locus_tag="OSJNBa0017E08.7"	
join(8942. .9423,10749. .10848,11202. .11535,11806. .11918, 12487. .12615,12796. .12889,12992. .13075,13315. .13946, 14712. .15001,15148. .15292,15926. .16023,16125. .16206, 16895. .16978)	
/locus_tag="OSJNBa0017E08.7"	
/codon_start=1	
/product="putative dentin phosphorin protein"	
/protein_id="AAP54439.1"	
/db_xref="GI:31432863"	
/translation="MIVRTYGRSRSPSCDGGGGRGGGSSQDAFEFDGEDDL VLISSQSSHPFAPSOESSMDFDPPPPRRRRRGGGGDYAPATAAAAAA TSUWAEYGGEMSVDEANFALDGLATGAPRRVRRASFLALGICAPRRVVRRAQ GLVQIIDAIVLNIIDDPCTIGAAALLFVLASDVQENHLLDSQVHFLKLLNPV NLVDSKAPSIGKLGISVKQMLGNSDKDCTSEILSKVEETLLSCOEIKSLDKDD KTTTRPELCPKWLALTMEKACISAVSVERTSDTVSVGGNFETLLSGDGLSDIFDV MDCCHSTLENLIKDTSTSDALDNEGTSLSQAALLKCLKILENATFLSDDNKTHLLM SRKLYPKRSLSFVGIISIELLSALSILQNSVSSSTPKSKSVQSCSCSADVMG GTSPNDGKRKNSKKKLLNSQTRHSLSSSEVSHITISSDAGSKQAFNCSPSIS SNAGSGSLGERHNGALKNIKKORGNANPIRGSSGWTISIRAHSDGNSREMAKRP RLSENVTIDSGGDDDFAFDVPQEPSPNELLGPKKSPQKHQDKSGNGVLVASHEPD QPEDLNQSGTTSLSFAKDESSLEDCLLASVKVLMNLANDNPSCGELIASCGGLNTWA	

SLIMKHFPSCFVVDNNTYNDVLDHSLSSONSKAHQVKKIOLRDHFLDPLVAIIG  
LLVNLVKDLSNRVLSARVVDLSQNPQSEETORDVIALLCVFLHSLQSGASESGT  
ISPDEESLMQGAERAEWIVEAYAAALLALSTESMKVRGNAISSCLFNNSUKILVPA  
LEKVSFHLQNLWITEETHSAVTEVIEKCLUS"  
/locus\_tag="OSJNBa0017E08.6"  
/note="similar to lysophosphatidic acid acyltransferase  
GB:AAF73736 GI:8163563 (Brassica napus); EST C28836 from  
this gene"  
join(19140..19391,19498..19683,19968..20003,20294..20700,  
21048..21149,21256..21483)  
/locus\_tag="OSJNBa0017E08.6"  
join(19316..19391,19498..19683,19968..20003,20294..20700,  
21048..21149,21256..21483)  
/locus\_tag="OSJNBa0017E08.6"  
/codon\_start=1  
/product="putative lysophosphatidic acid acyltransferase"  
/protein\_id="AAP54440.1"  
/db\_xref="GI:31432864"  
/translation="MTGLRPRPLAHAAGAGDATPSTAHAVVSGRGGRGVECOFHRV  
RRRQPOVAVATASRRRTTVRSDFDAAGAAATMGDSQALSVSTPTWKVMEYID  
VVSVRGVCFYAVTAIAIFLVAMVVHPLVLLDRYRRAQHYAKIWIATLISMIF  
YKLDVGENEIPNNSPVAHYANHOSFLDIYTLTLGRCKFKFSIKSIFMFPFIIGWAM  
YLLGPIPLRRMDSQLDKRCDLVKKGASVFFPEGTRSKDGLGAFKRGAFSVA  
TKTGAPVITPILTLGTGLMPGMEGILNSGVKLIHHPIEGNDAEKLCSARKVIAD  
TLILNGYGVH"  
/locus\_tag="OSJNBa0017E08.21"  
/note="similar to membrane associated protein Tb 291  
GB:CAA52141 GI:393394 (Trypanosoma brucei)"  
join(<25245..25732,25802..26164,26366..26506,26602..28303,  
28396..28542)  
/locus\_tag="OSJNBa0017E08.21"  
join(25245..25732,25802..26164,26366..26506,26602..28303,  
28396..28542)  
/locus\_tag="OSJNBa0017E08.21"  
/codon\_start=1  
/product="putative membrane-associated protein"  
/protein\_id="AAP54441.1"  
/db\_xref="GI:31432865"  
/translation="MARGSALLDGSVLPPSRIVSERQAGLPRRMPESATGREIVTLG  
EGRPADTPGRSFFLPFAMAGLVPPSSFFMDVLKFDYDLQMAHLTNAVNTLAFIAH  
LCMPFIVGRSLPRFMTFTVQSVSPVGGCYFQPMVPLNRYIPCLARKKWDWK  
SDCWRAPIGDGDDAVLDRLAGLSQGLTGAMVGYDLRRRIAPLORARCAWEYTG  
SEDTYRTHQVGRWQAPDEKFIYVQVRLNLSMEASLIPOGILPLCSDPDRASTLTIM  
TAVGASBERPGAASTQPEGERKKRLKMGETEPEGRNLISPPKWSFRRPSEIP  
SRPSHKSQSEADPRAAEARRRESDRRAADRLREAEAEAEAEAEAEAEAEAEAE  
EAAKQAE  
GDAQATSGAAGDEAGASLGTPSGDAQDQGLRDIPESGTISIGPSRVASSPRRLF  
PTPSIAPLSAPLPLQALAAANIAVLDSLQAEVAELDAWAEAEAEAEAEAEAEAE  
MVEVRKAHRHVSLEAEKVKVLAIEAEVEEERGAALIAIVSNEAODTLRLQYGSW  
EAEGLKDLAAQGVLDVAAREQARAGETEAAREEETLEAREAEAEAEAEAEAEAE  
DREAATVIREATLAHESACAEESALRHLHEDALITERAEAEAEAEAEAEAEAE  
EAE  
AEHTIADQGLTSSAGAEVLAELRAGEVPGMLWDVSRDLDRAGQVGLWRGRTVKA  
ANHGLAQRSLKMAQLRPLPELEKTIKSSRLDQAQVAVELVDAADHIVHSPGSA  
PRLAPAPNSDEEDNAGADDSDDEAGDPGVD"  
/locus\_tag="OSJNBa0017E08.20"  
/note="similar to gag-poi precursor from gypsy type  
retrotransposon GB:BAH84458 GI:5902445 (Oryza sativa)"  
complement(join(<29170..32047,32525..33441,35108..35540,  
35663..35940))  
/locus\_tag="OSJNBa0017E08.20"  
complement(join(29170..32047,32525..33441,35108..35540,  
35663..35940))  
/locus\_tag="OSJNBa0017E08.20"  
/codon\_start=1  
/product="putative gypsy-type retrotransposon"  
/protein\_id="AAP54442.1"  
/db\_xref="GI:31432866"  
/translation="MRMGQTKPSLRSEVTVPRCKGEAADPAPPRPDPAIPRLDLAWR  
RHEAARDGGGGSARCAVAARQARAGEVRNGMLPRARRRSVAACGRVRRASCMPWRPG

IGSRAAVVPLAASWAABRSOEGRGDOAKGVQVCGAGAGATYPALAOQMVVORELE  
SMGCLPRCSRSPFLELIPFFCEBFLGNIETAAQRQKRLPKOCHVPFGFLAKTG  
KAAGMNGVGLQSLGVLFLFSLMAQKAHDLSPVSGDDGSPNRRRRTPTPLP  
PQSPKGEALREVERSATSIAGDEERDGRLLVYDGSITPQALQAAGALLRH  
PPVVPDPSPARWLDDVAKLVMTVRQDLRAGRSSATKSGAAATGASRRARRA  
AAVHSAATSPSTPQKOLRGGPDARTSITERNRDSAAHATGASSSRVPOHGR  
GNQSPVPGVGCRAFPVASLRNVWPRFRPTIAEKYDGSVNPRAEMLANPVVVKAN  
GKLKBCIDYTLNACPKDPYLPRIIDQIVDSTAGRFISLGEKALPLKLRSGPP  
TWYKAEANALQLKAYLSSPPVLVAPDEPELLYLAATPQVVSAAALVEREDNSHP  
ALPHPTWPREGQGEAPENGGPRPTTGVPLPACTVGLGADPDGPEATVGRPK  
HLSPPDEANPLTRPGKEQEEAPENGGRLPLTTGVGLPACTTTCAPDQGPQK  
ATVGRPLLSDSPVISITEDKCAPRGCLDECPDAAPSEEDRPHKRVQRPVYVSEA  
LRDAKRVYQAKMLYAILMASRKLHYFOAHRVTVTYSYPLGOILNHRDNGRVVKW  
AIELSEFDLHPEPRHAIKSQLADFAVETAPPEVPSIPEASTDSQLPHTAHWQF

Query Match 51.3%; Score 987; DB 8; Length 300029;  
Best Local Similarity 79.5%; Pred. NO. 1.1e-159;  
Matches 1243; Conservative 0; Mismatches 185; Indels 135; Gaps 1;  
QY 292 GGTACATCAAGGAAGGAGTACCATAGCTGTGATAGGCTCAGTGTACACAGACACG 351  
DB 264381 GGTACATCAAGGAAGGAGTACCATAGGCTCAGTGTACACAGACACG 264322  
QY 352 TGGAGATCGTGAAAAAATCTCTCTTTCAGAGGAACTCTGGCCCGTCTATTCAAGAAC 411  
DB 264321 TGGAGATCGTGAGAAATCTCTCTTTCAGAGGAACTATGGCCCGTCTATTCAAGAAC 264262  
QY 412 TCATAGCTCAACCATATTGAAGAGGCTGCTGTTCTTAGTACCTGTATAGAAAGGAAA 471  
DB 264261 TCACAGCTCAATCATATTCAAGAGGCTGCTGTTCTTAGTACCTGTATAGAAAGGAAA 264202  
QY 472 TTTATGCTGCGCGCTATCATAGGAACCGTGTGTATCAGAGAAGTAGTGAGTGTGCA 531  
DB 264201 TCTATGCTGAGCTTTATTCGTGGAAACCGTGGGATTTAGAGAAGTGTGAGTGTGCA 264142  
QY 532 ----- 531  
DB 264141 AGGTAAGGATGTTTTCATCAAAATGCTCATTTGTTCTGTGCAATAGTCTCTTTCTTCGTCT 264082  
QY 532 ----- 531  
DB 264081 TTTCCCTTTTCTGTACTGAAGAAACGGGTTTGTCTTAGATGTTCTTAATTCCTCTGCT 264022  
QY 532 -----AGAAAGTGGTATTCGGCTTCGAGCTTAGGAGACCTGTTCA 576  
DB 264021 GCTTTTTTTTCAACAGAAAAGTGGAAATCCCTGCTCTCTGAGCTCAGGAGAGCATCTATTCA 263962  
QY 577 TCTTGGAAAGCAGTGATGCCACAGCCATCTGTTTGGAGTGTGCTGAGCTGCGCTTGAAGTCTT 636  
DB 263961 TGTTCGCTGACAGTATGCCACAGCCATCTGTTTGGAGTATCTGCTGGGCTTGAAGTCTT 263902  
QY 637 TGGTTCTCGGTGAAGGACAAATCTCTGCTCAGGTTTAAACAAAGTTGTGAGGAGTGCACAGA 696  
DB 263901 TGGTTCTCGGAGAGGACAAATCTCTGCTCAGGTTTAAACAAAGTTGTGAGGAGTGCACAGA 263842  
QY 697 ACAGTGAGGCTTGGGAAAGAACATCATGATAGGATGTTCAAGGATGCAATCATCTGCTGAAA 756  
DB 263841 ACAGTGAGGCTTGGGAAAGAACATCATGATAGGATGTTCAAGGATGCAATCATCTGCTGAAA 263782  
QY 757 AGCGTGTCCGAGGAGACCAACATATCATCTGCTGCTGCTGTTCTGTCAGTTTCAGGCGGG 816  
DB 263781 AGCGTGTCCGCTGGAGACTAAACATATCATCAGGTGCTGCTCTCTGTCAGTTTCAGTGCAG 263722  
QY 817 TTGAATCGGCTTGATGAAGCTTCGGAAGTCTGGAAGCACTGTGACGTAGGATGCTTCTGA 876  
DB 263721 TTGAATCGGCTTGATGAAGCTTCGGAAGTCTGGAAGCACTGTGACGTAGGATGCTTCTGA 263662  
QY 877 TTGCTGCTGTAAATGGGAAAGCTAGTATCAAAACATCTGTTTGGCAAGGATGCAAGA 936  
DB 263661 TTGCTGCTGCAAGATGGGAAAGTGTGTTGTTTAAACATTTGATTCGCAAGGATGCAAGA 263602  
QY 937 AGGTTGTTGTTGTAACCGCTCCGTGGAAAGGTTGATGCTATTCTGTCGAGGAGATGAAAG 996

263601 AAGTTGTTGGTGAACGGTTTCAGTGGAAAGGTTGATGCCATCCGCGAAGAGATGAAG 263542  
Db  
997 ATATAGAGATCGTGTACAGGCTCTCTCAGACATGTATCAAGCTGTGCTGAAGCTGATG 1056  
Qy  
263541 ACATTGAGATTGTGTACAGGCTCTCTACAGAGATGTATGAAGCGCTGCCAAGCTGATG 263482  
Db  
1057 TCGGTTCACACGACGCGCATCTGAACCTTCATTGTTGCGAAAGAACACACGACGAC 1116  
Qy  
263481 TCGGTTCACAAGCAGCGCATCCGAAACCCCATTTGTTACAAAGAGACGACGAGGCGC 263422  
Db  
1117 TCCCGCTGTCTCTGATATAATGGAGGTGTTTCGCTGTTTGTGACATATCTGTCCCA 1176  
Qy  
263421 TTCCCGCTATTCTGATGCTATGGTGTGTTTCGACTCTTTGTGACATATCCGTCCCA 263362  
Db  
1177 GGAATGTCAGCGCATGTGTCTGAAGTTGGCGTGCACGATGTACATGTCAGCAGCT 1236  
Qy  
263361 GAAATGTCAGCGCTGTGTGCTGAAGTTGGCCATGCGCGAGTATACAACTGATGACT 263302  
Db  
1237 TGAAGAGGTGTGGAAGCCCAAGAGGAGCGGCTCAGGAAAGCAATGAGGCGCAGA 1296  
Qy  
263301 TGAAGAGGTGTGGAAGCCCAAGAGGAGCGGCTTAGGAAAGCAATGAGGCGCCAA 263242  
Db  
1297 CAATCATCCGAAGAACTGAGACGTTTCGAGGATGAGGAGCTCGCTGGAGACCGTTC 1356  
Qy  
263241 CAATCATCCCAAGATTGAACCGTTTCGAGGATGAGGAGCTCGCTGGAGACTGTT 263182  
Db  
1357 CGACCATCAAGAGCTGAGTCTGACGCGGACAGATCAGGCGCTCGAGCTCGAGAGT 1416  
Qy  
263181 CGACTATCAAGAGCTGAGTCTGACGCGGACAGATCAGGCGCTCGAGCTGAGAGT 263122  
Db  
1417 GCCTGCAGAAAGTAGTGAGGACGCGCTCACAAAGAGATCAGGAGGACCATCGAGGAGC 1476  
Qy  
263121 GCCTCCAGAGATCGCGAAGAGCGCTCACAAAGAGATCAGGAGGACCATCGAGGAGC 263062  
Db  
1477 TGACACCGGATCGTTAAAGCTCTCCATGCGCGCTGCAGCACCTGAGTGCGAGC 1536  
Qy  
263061 TCAGCACCGGATCGTTAAAGCTCTCCATGCGCGCTGCAGCACCTGAGTGAGC 263002  
Db  
1537 GCAGCGACGCGGACCTTCAGCAGAGCTTCAGGAGATCAGGAGGACCATCGAGGAGT 1596  
Qy  
263001 GCAGCGACGCGGACCTTCAGCAGAGCTTCAGGAGATCAGGAGGACCATCGAGGAGT 262942  
Db  
1597 TCAGCCTCGACATGAGAGCGGATCATCGAGCAGAGATCAAGGCGCAAGGTGAGAGA 1656  
Qy  
262941 TCAGCCTCGACATGAGAGCGGATCATCGAGCAGAGATCAAGGCGCAAGGTGAGAGA 262882  
Db  
1657 CACAAACTGAGCGGAGAGCAATTTTCTACCAACCATATCTATATATAGCGTCTC 1716  
Qy  
262881 CCCAGAACTGAGATTGAAGAGAGATTTTTTTTTTTCAGCCCGGTATCTACTATGTATA 262822  
Db  
1717 CAA 1719  
Qy  
262821 CTA 262819  
Db  
RESULT 12  
AX081081  
LOCUS  
DEFINITION Sequence 11 from Patent WO0109304.  
ACCESSION AX081081  
VERSION AX081081.1 GI:13170004  
KEYWORDS  
SOURCE Glycine max (soybean)  
ORGANISM  
Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.  
1  
REFERENCE  
AUTHORS Cahoon, R.E., Gutteridge, S., Harvell, L.T., Rafalski, J.A., Tao, Y. and  
Weng, Z.  
TITLE Polynucleotides encoding aminolevulinic acid biosynthetic enzymes  
JOURNAL Patent: WO 0109304-A 11 08-FEB-2001;

E.I. DU PONT DE NEMOURS AND COMPANY (US)  
FEATURES Location/Qualifiers  
source 1..2055  
/organism="Glycine max"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3847"  
BASE COUNT 596 a 419 c 465 g 575 t  
ORIGIN  
Query Match 36.9%; Score 710.4; DB 6; Length 2055;  
Best Local Similarity 68.1%; Pred. No. 4e-112;  
Matches 1019; Conservative 0; Mismatches 471; Indels 6; Gaps 2;  
Qy 189 GGGGTGTCGGTTCGACGCGCGCGGTGGAGGCCAGGCGCAGGCGGTGGCAAGGCT 248  
Db 252 GGGGTATTTCGTCGACGCTCGACCCCTCTGATGCATCATCTGTGTCTCCAAATAATGCC 311  
Qy 249 GCCAGCGTCGCGCGCTCGAGCAGTTCAGAGTAT---CCGCGCAGCGGTACATGAAGAA 305  
Db 312 ACCGCTCTTCGCTCTTGAGCAGCTCAAGCTTCTGCGAGCTGATAGTATACAAGAA 371  
Qy 306 AGGATACCATAGCTGTGATAGGCTCAGTGTATACACAGCACGAGTGGAGATCGTGAA 365  
Db 372 AGAAGCAGCATTATCGCATTTGGGCTCAGTGTGCACACTGCACCTGTGGAATGCGTAA 431  
Qy 366 AAACCTGCTGTGACAGGAACTGTGCGCCGCTGCTATTCAAGAACTCAGTCTGAAAC 425  
Db 432 AAACCTGCCATTCCAGAGCAGAAATGGCTTAGAGCTATTGAGAGCTGTGTAGTCTGAAT 491  
Qy 426 CATATTCAAGAGGCTGTGTTTCTTAGTACCTGTAATAGAAATGGAATTTATGTGTGGCG 485  
Db 492 CATATTGAAGAGCAGCTGTCTGAGTACCTGCAATCGAATGGAGATATATGTTCTGCC 551  
Qy 486 CTATCATGGAACCGTGTATCAGAGAAAGTAGTGGACTGGATGTCGAGAAAGTGGTATT 545  
Db 552 CTGTCCCAACATCGTGTGTCAAAGAGTCAATGAAATGATGTCAGAAACAAGTCTGTC 611  
Qy 546 CCGCTTCCGAGCTTAGGAGCAGCTGTTTCTGTTGAGTCTCGGTGAAGGACAAATCTTCTCT 665  
Db 612 CCGTTCAGAGCTTAGCCAGCACCGGTTTTTACTTTTACAAATGATGTCACACAGCAT 671  
Qy 606 CTGTTTCAGGTGTGAGTGGCTTGAATCTTTGTTCTCGGTGAAGGACAAATCTTCTCT 665  
Db 672 CTTTTCAGAGTATCAGCAGCTTGTGACTCTCTGTTTGGGGGAAGGTCAATCTTCTCT 731  
Qy 666 CAGGTTAAACAAGTTGTGAGGAGTGGACAGAACAGTGGAGCTTTGGGAAGAACATCGAT 725  
Db 732 CAGGTTAAGCAAGTTGTTAAAGTTGGACAAGGAGTTAAACGCTTTGGGAGAAATATCAGT 791  
Qy 726 AGGATGTTCAAGGATGCAATCAGCTGGAAGAGGCTGTCGACGAGAGACCAATATCA 785  
Db 792 GGGCTATTCAAGCATGCAATTTACTGTCGGGAAAGGGTTAGAACTGAGACTAATATTGCT 851  
Qy 786 TCTGCTGCTGTTTCTGTCAGTTAGCGCGGCTTGAACCTGGCCCTGATGAAGCTTCCGAAG 845  
Db 852 TCTGGGCGAGTTCTGTGAGCTCAGCTGCGGTTGAGTTGAGCTATATGAGTTACTCTGAA 911  
Qy 846 TCTGAAGCAGCTGTCAGCTAGGATGCTTCTGATTTGCTGCTGTTAAAATGGGAAAGCTAGT 905  
Db 912 GCCTCACACGATTAATGCCAGATGTTGGTTATTGGTGTCTGCGCAAGATGGGAAAGCTGTG 971  
Qy 906 ATCAACATCTGTTGTCGAAGGATGCAAGAGGTTGTTGTTGTTGAAACCGCTCCGTGGAA 965  
Db 972 ATCAACATTTGGTGGCAAGAGGTTGCAAAAGATGTTGGTTGTTGCAATAGAACTGAGGAG 1031  
Qy 966 AGGCTGATGCTATTTCGTGAGGAGATGAAAGATATAGAGATCGTGTACAGGCTCTCTCA 1025  
Db 1032 AGAGTTGCTGCATACGTGGAAGAACTGAGAGATATTGAGATTATCTACAAACCCCTTCA 1091  
Qy 1026 GACATGATCAAGCTGCTGCTGAAGCTGATGTGCTGTTTCACGACCGCATCTGAAACT 1085  
Db 1092 GAAATGCTCACCTGTGTCGCGAAGCAGATTTAGTTTTTACCAGTACTGCTGATCAGAAAC 1151

1086	Qy	TCATTGTTGCGAAAAGAACCGCAGAGGCACTCCCCCTGCTCTGTGATCTATGGGAGGT	1145
1152	Db	CCATTATTCTTGAAGAACAATGTCAAGGACCTTCTCTCGAAGTCAAGAAGTTGGAGGC	1211
1146	Qy	GTTCCGCTGTTTGTGCAATATCTGTCCCGAGGAATCTCAGCCATGTGTCTGAAGTT	1205
1212	Db	CGTCCGCTTTTTCATTGATATCTCTGTTCCCGGAATGTGGGTTCAATGTCTCAGACCTT	1271
1206	Qy	GGCGCTGCA CGAGTGTCAATGTGCGAGCACTGTAAAGAGGTGGTGAAGCCAAACAAGGAG	1265
1272	Db	GAGTCTGTGCGAGTTTACAATGTTGACGACCTTAAAGAGGTGTGGCTGCCAATAAGAG	1331
1266	Qy	GACCGCTCAGGAAGCAATGGAGCCGCGACAAATCATACCGAAGAACTGAGACGGTTC	1325
1332	Db	GATCGCTTAAGAAAGCAATGGGAAGCAGACGCAATCATTCCTGAAGAATCTAAGACCAATC	1391
1326	Qy	GAGCATGAGGAGACTCGCTGGAGACCGTTCGAGCAATCAAGAACTGAGGTCTGACGG	1385
1392	Db	GAAGCTTGGAGGACTCACTGGAAACTGTGTTCTACTATTGAATAATTGAGGGCTTTATGCT	1451
1386	Qy	GACAGGATCAGGGCCTCGSAGCTCGAGAAGTGCTGCAGAAAGTAGTGTAGGACGCCCC	1445
1452	Db	GAAGAATCAGGCTTGTCTGAGCTTGAGAAGTGCTTAGTAAAGTGGTGATGA--TATA	1508
1446	Qy	ACCAAGAATGAGGAGAGCCATCGAGGAGCTGAGCAACCGGCAATCGTTAAACAAGTCTCT	1505
1509	Db	CCAAAGAAAACGGGAGAGCTGTGGATGACCTTAGTCCGGGTTATAGTGAATAAGATTGCTT	1568
1506	Qy	CATGGCCCGCTCGAGCACTCTGAGTGGCGGCGGAGCGACCGCACCTTGCAGGACG	1565
1569	Db	CATGGTCCAATGCAACATTTAAGGTGTGATGGGAACGACAGCCGGACTCTTAGTGAGACA	1628
1566	Qy	CTCGAAGAACATGCAACGCCCTCAACCGGATGTTTCAGCCTCGACATGGAGAAGCGCATC	1625
1629	Db	CTGAGAACAATGAATGCTTTTGAATAGGATGTTCAACCTTGAGACAGAAATATCTGTTTG	1688
1626	Qy	GAGCAGAGAGTCAAGCCCAAGTGTGAGAGACACAAAACCTGAGGCCAGGAAGCAAT	1681
1689	Db	GAGGAGAAGATCGAGCAAGGTGCAACAAACCCAGAAATGAAATCTAACACCAAT	1744

**RESULT 13**

D50407	D50407	2072 bp	mRNA	linear	PLN 10-FEB-1999
LOCUS	Cucumber mRNA for glutamyl-tRNA	reductase, isozyme 1	complete cds.		
DEFINITION	D50407				
ACCESSION	D50407				
VERSION	D50407.1	GI:1694925			
KEYWORDS	hemA; glutamyl-tRNA reductase.				
SOURCE	Cucumis sativus (cucumber)				
ORGANISM	Cucumis sativus				

REFERENCE  
AUTHORS

**TITLE**  
Differential expression of two hema mRNAs encoding glutamyl-tRNA reductase proteins in greening cucumber seedlings

**JOURNAL**  
Plant Physiol. 110 (4), 1223-1230 (1996)

**PUBMED  
REFERENCE  
AUTHORS**

**TITLE**  
JOURNAL  
Direct Submission  
Submitted (21-APR-1995) Ryouichi Tanaka, Kyoto University,  
Department of Botany, Faculty of Science, Sakyo-ku,  
Kitashirakawa-oiwakecho, Kyoto, Kyoto 606-01, Japan  
(E-mail: ryoueok-lab.bot.kyoto-u.ac.jp, Tel: 075-753-4147,  
Fax: 075-753-4257)

## FEATURES

```

source
1. .2072
/organism="Cucumis sativus"
/mol type="mRNA"

```

gene	CDS
g00001	g00001
g00002	g00002
g00003	g00003
g00004	g00004
g00005	g00005
g00006	g00006
g00007	g00007
g00008	g00008
g00009	g00009
g00010	g00010
g00011	g00011
g00012	g00012
g00013	g00013
g00014	g00014
g00015	g00015
g00016	g00016
g00017	g00017
g00018	g00018
g00019	g00019
g00020	g00020
g00021	g00021
g00022	g00022
g00023	g00023
g00024	g00024
g00025	g00025
g00026	g00026
g00027	g00027
g00028	g00028
g00029	g00029
g00030	g00030
g00031	g00031
g00032	g00032
g00033	g00033
g00034	g00034
g00035	g00035
g00036	g00036
g00037	g00037
g00038	g00038
g00039	g00039
g00040	g00040
g00041	g00041
g00042	g00042
g00043	g00043
g00044	g00044
g00045	g00045
g00046	g00046
g00047	g00047
g00048	g00048
g00049	g00049
g00050	g00050
g00051	g00051
g00052	g00052
g00053	g00053
g00054	g00054
g00055	g00055
g00056	g00056
g00057	g00057
g00058	g00058
g00059	g00059
g00060	g00060
g00061	g00061
g00062	g00062
g00063	g00063
g00064	g00064
g00065	g00065
g00066	g00066
g00067	g00067
g00068	g00068
g00069	g00069
g00070	g00070
g00071	g00071
g00072	g00072
g00073	g00073
g00074	g00074
g00075	g00075
g00076	g00076
g00077	g00077
g00078	g00078
g00079	g00079
g00080	g00080
g00081	g00081
g00082	g00082
g00083	g00083
g00084	g00084
g00085	g00085
g00086	g00086
g00087	g00087
g00088	g00088
g00089	g00089
g00090	g00090
g00091	g00091
g00092	g00092
g00093	g00093
g00094	g00094
g00095	g00095
g00096	g00096
g00097	g00097
g00098	g00098
g00099	g00099
g00100	g00100

**BASE COUNTRY OF ORIGIN**

Query Match	36.7%	Score 706.2;	DB 8;	Length 2072;
Best Local Similarity	68.8%	Pred. No. 2.1e-111;		
Matches 1001; Conservative	0;	Mismatches 448;	Indels	6;
			Gaps	2;

Qy	241	CAAAAGGCTGCCAGCTGGCGGCGCTCGAGCAGTTCAAGATATCCGC- --CGACCGGTACA	297
Db	310	CTAAATCTTCTAATCTCTCTGCCCTTGAAACAGCTCAAGACCTCTGCTGTTGCAGATATA	369
Qy	298	TGAAGGAAGGAGTACCATAGCTGTCATAGGCTCAGTGTACACACAGCAGCAGTGGAGA	357
Db	370	CAAAAGGAAGACAGCATTTGTTGTCATTTGGGCTTAGTATTATCAACACCTGTTGAAA	429
Qy	358	TGCGTGAAAAAATTCGTGTTTCAGAGGAACCTGCGCCCGTCTATTCAAGAACCTCACTA	417
Db	430	TGAGAGAAAAAATCGGCCAATTCCTGAGGCAGAGTGCGCTCGAGCCATTGGAGAGCTCTGTG	489
Qy	418	GCCTGAACCATATTGAAGAGGCTGCTGTTCTTAGTACCTGTTAATAGAAATGGAAAATTTATG	477
Db	490	GCTTTAAACCACTTGAAGAAGCAGCGGTTCTTAGCACCTCGCAACAGAAATGGAGATATATG	549
Qy	478	TGCTGGCGCTATCATTCGAAACCGTGGTATCAGAGAAGTAGTGACCTGGATGTCGAAAGAAA	537
Db	550	TTGTTGCTTTGTTCTCAGCATCTGTGNGTGAAGGAAGTGACTGAAATGGAATGTTCCAAACAA	609
Qy	538	GTGGTATTCGCGCTTCGAGAGCTTAGGAGACA CTTGTTCACTTTGCGAAGCAGTGTATGCCA	597
Db	610	GTGGAAATCCCGGTTTCGGAGATTTTGCACGACCGGTTTTTGCTGTATTAACAACGATGCCA	669
Qy	598	CAGCCCATCTGTTTGAGGTGT CAGCTGGCTTTGACTCTTTGGTCTCTCGTGAAGCAGCAA	657
Db	670	CACAAATATTTTCGAAGTTTCAGCAGGCTTGACTCTCTTGTTCTTAGGAAGGCGCAA	729
Qy	658	TCCTTGCTCAGGTTAAACAAGTTGTGAGGAGTGGACAGAA CAGTGGAGGCTTTGGGAAAAGA	717
Db	730	TCCTTGCTCAGGTTAAACAAGTTGTCAAGTTTGGACAGGGTGTTCAGGATTTTCGAAGGA	789
Qy	718	ACATCGATAGGATGTTCAAGGATGCAATCACTGCTGGAAAGGGTGTCCGACGCGAGACA	777
Db	790	ACATTAGTGGGCTTTTCAAGCATGCAATCA CCGTGGGTAAACGGGTAAAGACAGAGACTA	849
Qy	778	ACATATCATCTGGTGTCTGTTCTGT CAGTT CAGCGCGCGGTTGAACCTGGCCCTGATGAAGC	837
Db	850	ATATTGCTGCTGAGCTGTTTCTGTGAGTTTCGCTGCTGT CGAATTTGGCCTTGATGAAGC	909
Qy	838	TTCCCGAAGTCTGAAGCATGT CAGCTAGGATGCTTTCGATTTGGTGCTGGTAAAATGGGAA	897
Db	910	TTCCCTGAACCTTCA CATGCCATCGCCAGATGTTGGTCAATTCGAGCTGGTAAAGATGGGAA	969

Qy	898	AGCTAGTGTCAAAACATCTCGTTGCTCCAAAGATGCAAGAAGTTGTTGTTGGTGAACCGCT	957
Db	970	AGCTTGTAAATTAACATTTGTTGCTAAAGCGCTGCAAAAAGATGGTTGTTGTGAATCGT	1029
Qy	958	CCGTGGAAAGGGTGGATGCTATTTCGTGAGAGATGAAAGATATAGAGATCGTGTACAGGC	1017
Db	1030	CTGAAGAGAGAGTCAACAGCCATCCGTGAGGAGATGAAGGATGTCGAGATAATCTTACAAGC	1089
Qy	1018	CTCTCTCAGACATGTATCAAGCTGCTGCTGAAAGCTGATGTGTCGTGTTCCACGAGCACCGCAT	1077
Db	1090	CCCTCAGCGAAAATGCTTAGCTGTACACTGAAGCAGACGTTGATATTATCCAGCACAGCTT	1149
Qy	1078	CTGAAACTTCATTGTTTCGCAAAAGAACACGAGAGGCACCTCCGCCCTGTCTCTGTACTTA	1137
Db	1150	CTGAAAGTCTTTTGTGTTTACAAAGAGAGAGTTAAGGATCTTCCCCCTGTTTGGCATGATG	1209
Qy	1138	TGGAGAGTGTTCGCGCTGTTGTGCACATATCTGTGCCACGAATGTCCAGCGCATGTGTGT	1197
Db	1210	TTGGGGCTTGAGGCTTTTCATCGATATCTCTGTTCTCGAAGCTCGGAGCATGTATCA	1269
Qy	1198	CTGAAGTTGGCGTGCACAGAGTGTACAATGTTCGACGACTTTGAAAGAGGTTGGTGCAGGCCA	1257
Db	1270	ACAACTCTCGAAGATGTACGAGTGTACAATGTTGACGACCTTAAGGAGGTAGTTGCTGCAA	1329
Qy	1258	ACAGGAGGACCGCTCAGGAAAGCAATGAGGCGCAGACAATCATCATCCACGGAAGACTGA	1317
Db	1330	ACAAAGGAAGATAGGCTCCGAAAAGCAATGGAAGCACAAATCAATTTATCACCTGGAAGAAATCCA	1389
Qy	1318	GACGGTTCGAGGATGAGGAGACTCGCTGGAGACCGTTCCGACCATCAAGAAGCTGAGGT	1377
Db	1390	AGCAATTTGAAGCTTGGAGGGATTCATTGAGACTGTTTCCAACTTTAAAAACTTAGGG	1449
Qy	1378	CGTACGGCGACAGATCAGGGCTCGAGCTTCGAGAAAGTCCCTGCAGAAAAGTAGGTGAGG	1437
Db	1450	CTTATGCGGAAAAGAAATCAGAACTCCCGAGTTAGAAAAATGCCTATCAAGATCGGTGATG	1509
Qy	1438	ACGCCCTCACCAAGAAGATGAGGAGAGCCATCGAGGAGCTGACACCGGCATCGTTAACA	1497
Db	1510	A---TATCCCAAGAGACTCGACGACTGTGGATGATCTTAGTCGTGGTATAGTTAACA	1566
Qy	1498	AGCTCTTCATGGCCCGCTCGACACCTGAGGTGCGACGGCAGCGACCGCACCCCTTG	1557
Db	1567	AGCTGCTTCATGGTCCCTATGCAGCATTTGATGCGACGGGAGTGACGCGAACTTTAA	1626
Qy	1558	ACGAGACGCTCGAGAACATGCAAGCCCTTCAACCGGATGTTTACGCTTCGACATGGAGAAG	1617
Db	1627	GTGAAACTTTGGGAAATGACCGCTCTAAATAGAATGTTTTCAGCTTGGAAAACAGAAATCG	1686
Qy	1618	CGATCATCGAGCAGAAGATCAAGGCCAAGGTGGAGAGACACAAAACCTGAGGCCAGGAAG	1677
Db	1687	CAGTGTGGAACAGAAAATCAGACTAAAGTGGAAACAAAACCAAGAAAGTAGAGTAGTAA	1746
Qy	1678	CAATTTTCTTACCAC	1692
Db	1747	TATTATATGAAGCTC	1761

RESULT	14
AX506506	
LOCUS	AX506506
DEFINITION	Sequence 1201 from Patent WO0218655.
ACCESSION	AX506506
VERSION	AX506506.1
KEYWORDS	GI:23387743
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
REFERENCE	1 Harper, J.F., Kreps, J.J., Wang, X. and Zhu, T. Stress-regulated genes of plants, transgenic plants containing same, and methods of use
AUTHORS	
TITLE	

JOURNAL		Patent: WO 0216655-A 1201 28-FEB-2002; The Scripps Research Institute (US) ; Syngenta Participations AG (CH)	
FEATURES	source	Location/Qualifiers	
		1. .1632 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702"	
BASE COUNT	463 a	-289 c	452 t
ORIGIN			
	Query Match	35.3%;	Score 679.6; DB 6; Length 1632;
	Best Local Similarity	68.6%;	Pred. No. 7.9e-107;
	Matches 985;	Conservative 0;	Mismatches 439; Indels 12; Gaps 3;
Qy	235	CCGTGGCAAAAGGCTGCCAGGCTCGCCGCTCGAGCAGTTCAAGATAT--	CCGCGCAGCC 291
Db	206	CTGGCTTAATGACGTAGCAATCTGCTCTTGAACTCAAGAACTCGCAGCTGATC	265
Qy	292	GGTACATGAAGAAAGGAGTACCATAGCTGTGATAGGCCTCAGTGTACACAGCACCAG	351
Db	266	GATATACAAAGAAAGACAGTATTGTTGATTGGACTTTAGTATTCACACAGCTCCTG	325
Qy	352	TGAGATGCGTGAAGAACTTGCTGTTGACAGAGAACTGTGGCCCGCTGCTATTCAAGAAC	411
Db	326	TTGAGATGCGTGAAGAGCTTGCTATACCTGAAGCTGAATGGCCCAAGAGCTATTGCTGAAT	385
Qy	412	TCACTAGCTGAACCATATTGAAGAGGCTGCTGTTCTAGTACTCTTAATAGATCGAAA	471
Db	386	TGTGTGTTTGAATCATATTGAAGAAGCTGCTGTGCTTAGTACTTGTAAACCGTATGGAGA	445
Qy	472	TTTATGTGTGGCGCTATCATGAAACCGTGGTATCAGAAAGTAGTGGACTGATGTCGA	531
Db	446	TTTATGTTTATAGCTTTATCTCAGCATCGTGGAGTTAAGAGTTACTGATGATGTCAA	505
Qy	532	AGAAAAGTGTATTCCCGCTTCCGAGCTTAGGAGCACTGTTTCATCTTTCGGAAGCAGTG	591
Db	506	AGACGAGTGGGATCCCAAGTTTCAGAGATTGTCAACACCGCTTTCTGTTGTCAACAAGG	565
Qy	592	ATGCCACAGCCATCTGTTTGAAGTGTGAGTGTGAGCTGGCTTGACTCTTTGTTCTCGGTGAAG	651
Db	566	ATGCTACACAGCATATTTTGAAGTCTCAGCTGGTCTTGACTCTCTTGTCTTAGSAGAAG	625
Qy	652	GACAAATCTTGTCTCAGGTTAAACAAGTTGTGAGGAGTGGACAGACAGTGGAGGCTTG	711
Db	626	GTGAGATCTTTGACAGGTGCAACAAGTTGTGAAAGTTGTTCAAGAGTGAATGGCTTTC	685
Qy	712	GAAAGAACATCGATAGATGTTTCAAGGATGCAATCACTGCTGGAAGACGTTGCCGAGCG	771
Db	686	GGAGGAATATCAGCGGCTGTTTAAACACCGGATACTGTTGGTAAAGCGTGTGTAGAACAG	745
Qy	772	AGACCAACATATCATCTGGTGTGTTTCTGTGAGTTTCAGGCGGTTGAACTGGCCCTGA	831
Db	746	AGAAATATTGCTTCTGGTGGGTTTTCCGGTAGCTCAGCTGCTGTTTGAACCTTGCTCTGA	805
Qy	832	TGAAGCTTCCGAAGTCTGAAGCACTGCTCAGCTAGGATGCTTCTGATTTGGTCTGGTAAAA	891
Db	806	TGAAGCTTCCGAGTCTTCAACAGTATCTGCTAGGATGTGTAAATGGCGCTGGAAAAA	865
Qy	892	TGGGAAGCTAGTGATCAAAATCTGCGTTGCCAAGGATGCAAGAGGTTGTTTGGTGA	951
Db	866	TGGGAAGCTTGTGATTAAGCATTTGATGGCTTAAGGTTTGCAAAAAGTGGTGTAGTCA	925
Qy	952	ACCGCTCCGTGGAAAGGTTGGATGCTATTTCGTGAGGAGATGAAGATATAGAGATCGTGT	1011
Db	926	ACAGAAGTGAAGAGAGGGTTTCAGCTATCCGTGAGGAGATGCTCGTATTTGAGATTTAT	985
Qy	1012	ACAGGCTCTCTCAGACATGTATCAAGCTGCTGCTGAAGCTGATGTCGTTTCACAGCA	1071
Db	986	ATCGACTCTTGATGAGATGCTAGCTTTGCTTTCAGAAGCGGATGTTGTTTACAGCA	1045
Qy	1072	CCGCATCTGAACCTTCATTGTTGCGAAAAAGAACACGACAGAGCACTCCCGCTGCTCTCTG	1131

Db	1046	CAGCCTCTGAGACGCCATTGTTCTTTAAAGGAGCATGTAGAGAAATCTCCCTCAAGCTTCTC	1105
Qy	1132	ATACTATGGAGGTGTTCGCTGTTTGTTCGACATATCTGTCGCCAGGAATGTCAGCGCAT	1191
Db	1106	CAGAGTTGGAGGACTGAGGCAATTTTGTGTGACATCTCTGTTCCGGAATGTTGATCTT	1165
Qy	1192	GTGTCTCTGAAGTTGGCCCTGACAGGTGTCAATGTTCGACGACTTGAAGAGGTGGTGG	1251
Db	1166	GTGTTGGTGAAGTAGAACAACAGCAGCGGTTTCAATGTGGACGATCTCAAGAGGTTGTTG	1225
Qy	1252	AAGCAACAAGGAGACCGGCTCAGGAAGCAATGAGCGCGGAGCAATCATCAGCGAAG	1311
Db	1226	CTGCCAACAAGAGACAGGAGTGAAGAAAGCAATGGAAGCTCAACCAATAATCAGAGG	1285
Qy	1312	AAGTGAAGCGTTTCAGGAGCATGAGGAGGACTGCTCGAGACGTTCCGACCATCAAGAAGC	1371
Db	1286	AATCCACAGCTTTGAAGCATGGAGGAGTTCATTTGAGACGGTTCCACGATCAAGAAGT	1345
Qy	1372	TGAGTCTGACCGGACAGGATCAGGGCTCGAGCTCGAGAAGTGCCTGCGAGAAGTAG	1431
Db	1346	TAAGAGCTTATGACAGAGAAATCAGATGGCAGAGCTCGAGAAGTGCATGTCCAAAATGG	1405
Qy	1432	GTGAGGACGCTTACCAAGAGATGAGAGAGCATCGAGAGCTGAGACCGGATCG	1491
Db	1406	GAGATGA---TATCAACAAGAAACAACAGAGCGGTTGATGACTTAAGCCGAGGTATCG	1462
Qy	1492	TTAAACAAGCTCTCCATCGCCGCTGCGAGCACTGAGGTGCGAGCGGACGACGCCGA	1551
Db	1463	TGAACAGATTTCTACATGGTCCATGCGAGCTTTGAGATGTGACGGAAGTGCAGTAGAA	1522
Qy	1552	CCCTTGACGAGCGCTCGAAGCAATGACGCCCTCAACCGGATGTTACGCTCGACATGG	1611
Db	1523	CGCTGAGCGAGACCTTTCGAGAATATGCAATGCTTTTGAACAGAATGTACGGTC----	1576
Qy	1612	AGAGGCGATCATGAGCAGAGATCAGGCGCAAGTGGAGGAGAGACACAAAACCTGA	1667
Db	1577	AGAAAGACATTTTGAGGAAAGCTTAAGGCAATGGCGGACAAACAACAGAGTA	1632

RESULT 15	AY096600	1663 bp	mRNA	linear	PLN 18-SEP-2002
LOCUS	AY096600				
DEFINITION	Arabidopsis thaliana putative glutamyl-tRNA reductase (At1g58290)				
ACCESSION	AY096600				
VERSION	AY096600.1				
KEYWORDS	FLI CDNA.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.				
	Arabidopsis Open Reading Frame (ORF) Clones				
	Unpublished				
TITLE	Arabidopsis				
JOURNAL	2 (bases 1 to 1663)				
REFERENCE	Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.				
AUTHORS	Direct Submission				

JOURNAL	Submitted (16-APR-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA				
COMMENT	The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.				
	The Salk, Stanford, PGSC (SSP) Consortium members constructed and sequenced the PUN1 (ORF) clones using the RAPL cDNAs: Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.				
	Yamada, K. (SSP/PGSC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGSC) contributed equally to this work as PIs.				
	Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.				
FEATURES	Location/Qualifiers				
source	1. .1663				
	/organism="Arabidopsis thaliana"				
	/mol_type="mRNA"				
	/db_xref="taxon:3702"				
	/chromosome="1"				
	/clone="U19478"				
	/note="This clone is in PUN1 51."				
gene	ecotype: Columbia				
	1. .1663				
	/gene="At1g58290"				
CDS	1. .1632				
	/gene="At1g58290"				
	/codon_start=1				
	/evidence=experimental				
	/product="putative glutamyl-tRNA reductase"				
	/protein_id="AA02050.1"				
	/db_xref="GI:20465535"				
	/translation="MAVSAFVGPCKLETLLNHNHSPSSSSSSSVSTPGLNGVRV LPKNRTFRGLQKARCELSASNAASISALEOLKNSAADRYTKERSIVVIGL SIHTAPVEMREKLAIPAEWPRALAECLGNHIEAAVLSTCNREIIVLALSQRHV KEVTEWMSKTSIGIPVSEICORHFLYINKDATOHIFVSAGLSDVLGEGQLAQKQV VKVQGVNGFRNIGLFIHATITVGEKRTETNIFASGVSSAAVELALMKLPQSN VSARMCVIGAGKMGKLVIKHLMAGCTKVVVNRSERVSAIREMPCIEIIVRPLDE MLACSEADVFTSTASETPLFLKEHVNLPQASPEVGLRHFVDISVPRNVGSCVGE VETARVTVNDDLKEVVAANKEDRMKAMEAQTIIESTQFEAWRDSLETPTTKLR AYAEIRVASELEKCMKMGDDINKKTTTAVDDLSRGIVNRFPHGPMOHLRCDGSDSRT LSETLENHNLNRMYGLEKDLILEKUKAMAEQQOK"				
3'UTR	1633. .1663				
	/gene="At1g58290"				
BASE COUNT	477 a	292 c	433 g	461 t	
ORIGIN					
Query Match	35.3%; Score 679.6; DB 8; Length 1663;				
Best Local Similarity	68.6%; Pred. No. 7.9e-107;				
Matches	985;	Conservative	0;	Mismatches	439;
				Indels	12;
				Gaps	3;
Qy	235	CCGTGGCAAGGCTGCCAGGCTCGCGCCCTCGAGCAGTTCAAGATAT---	CCGCCGACC	291	
Db	206	CTGGCTTAATGAGCTAGCATCTCTGCTTTGAACAACTCAAGACTGTCAGCTGATC	265		
Qy	292	GGTACATGAAGAAAGAGTACCATAGCTGTGATAGGCTTCAGTGTACACACAGCAGCAG	351		
Db	266	GATATACAAAGAAAGAGCAGTATTGTTGTTGATTTGACTTAGTATTTCACACAGCTCTG	325		
Qy	352	TGAGATGCGTGAAGAACTTCTGTTGTCAGAGGAACTGTGGCCCGGCTGCTTCAAGAAC	411		
Db	326	TTGAGATGCGTGAAGAGCTTGTCTATACCTGAAGCTGAATGGCCCAAGAGCTATTGCTGAAT	385		
Qy	412	TCAGCTGACCTGACCATATTGAAGAGGCTGCTGTTCTTAGTACCTGTAATAGAAATGAAA	471		



Db 386 TGTGTGGTTTGAATCATATTGAAGAAGCTGTGTGCTTAGTACTTGTAAACCGTATGGAGA 445  
Qy 472 TTTATGTGGTGGCGCTATCATGGAACCGTGTGTATCAGAGAAAGTAGTGACTGGATGTGCA 531  
Db 446 TTTATGTGTTTAGCTTTATCTCAGCATCGTGGANGTTAAAGAAGTTACTGAAATGGATGTCAA 505  
Qy 532 AGAAAAGTGGTATTCCCGCTCCGAGCTTACGAGCTTAGGAGACCTGTTTCATCTTTGGCAAGAGTG 591  
Db 506 AGACGAGTGGGATCCAGTTTCAGAGATTTGTCAACACCCGTTTCTGTGTGTACAAACAGG 565  
Qy 592 ATGCCACACGCGCATCTGTTTGAAGTGTGAGTGGGCTTGAATCTTTGGTGTCTCCGTGAAG 651  
Db 566 ATGCTACACAGCATATATTTGAAGTCTCAGCTGGTCTTGACTCTCTTGTCTTAGAGAGG 625  
Qy 652 GACAAATCCTTGTCTCAGGTTAAACAAAGTTGTGAGAGTGGACAGAAACAGTAGTGAGGCTTGG 711  
Db 626 GTCAGATCCTTGCAACAGGTGAACAAAGTTGTGAAAGTTGGTCAAGGAGTGAATGGCTTTG 685  
Qy 712 GAAGAACATCGATAGGATGTTCAAGGATGCAATCACTGCTGGAAGCGTGTCCGACGG 771  
Db 686 GGAGGAATATCAGCGGGCTGTTTAAACACCGGATAACTGTTGGTAAGCGTGTAGAACAG 745  
Qy 772 AGACCAACATATCATCTCGTGTCTTCTGTGAGTTTCAAGCGGCTTGAACCTGGGCCCTGA 831  
Db 746 AGACAAATATTGCTTCTGTTGGGTTTCGGTTAGCTCAGCTGCTGTTGAACCTGCTCTGA 805  
Qy 832 TGAAGCTTCCGAAGTCTGAAGCATGTCAGCTAGGATGCTTCTGATTGGTGTCTGGTAAAA 891  
Db 806 TGAAGCTTCCGAGTCTTCAAACGTTATCTGTAGGATGTGTGTAATTTGGCGCTGGAAAA 865  
Qy 892 TGGGAAGCTAGTGATCAACATCTGTTGCCAAAGGATGCAAGNAGSTTGTGTGGTGA 951  
Db 866 TGGGAAGCTTGTGATTAAAGCATTTGATGGCTTAAGGTTGCAACAAAGTGTGTGTAGTCA 925  
Qy 952 ACCGCTCCGTGAAAGGCTGATGCTATTCTGTGAGGATGAAGATATAGAGATCGTGT 1011  
Db 926 ACAGAAGTGAAGAGAGGTTTCAGCTATCCGTGAGGAGATGCCCTGTTATTGAGATTAT 985  
Qy 1012 ACAGGCTCTCTCAGACATGATCAAGCTGTGCTGTAAGCTGATGTCTGTTCACCGCA 1071  
Db 986 ATCGACCTCTTGATGAGATGCTAGCTTGTGTCTCAGAAAGCGGATGTTGTGTTTACCAGCA 1045  
Qy 1072 CGCATCTGAAACTTCTATTGTCGAAAGAACCGCAGAGCACTCCCGCTGCTCTG 1131  
Db 1046 CAGCTCTGAGCGCATTTGTTCTTAAAGGAGCATGTAGAGAACTCTCCCTCAAGCTTCTC 1105  
Qy 1132 ATACTATGGAGGTTTCCGCTGTTTGTGACATATCTGTCCCGAGGAATGTCAAGCGCAT 1191  
Db 1106 CAGAGTTTGGAGGACTGAGGCAATTTTGTGACATCTCTGTTCCGAGAAATGTTGGATCTT 1165  
Qy 1192 GTGTGCTGTAAGTTGGCGCTGCAAGAGTGTACAATGTGACGACTTGAAGAGGTGGTGG 1251  
Db 1166 GTGTGTGTAAGTAGAACAAGCAGCGGGTTTACAATGTGGAGCATCTCAAAGAAGTTGTTG 1225  
Qy 1252 AAGCCACAAGAGAGACCGGCTCAGAAAGCAATGGAGGCGCAGACATCATCACCGAAG 1311  
Db 1226 CTGCCACAAGAGACAGGATGAGAAAGCAATGGAAGCTCAAACCATATTCAGAGG 1285  
Qy 1312 AACTGAGACGTTTCAGGCAATGGAGGACTCGCTGGAGACCGTTCCGACCATCAAGAAGC 1371  
Db 1286 AATCCACACAGTTTGAAGCATGGAGGGATTCAATGGAGACGGTTCCAAACGATCAAGAGT 1345  
Qy 1372 TGAGGTCGTACGCGACAGGATCAGGCGCTCGGAGCTCGAAGTGCCTGCGAGAAAGTAG 1431  
Db 1346 TAAGAGCTTATGCAAGAGAAATCAGAGTGGCAGAGCTCGAGAAGTGCAATGTCCAAATGG 1405  
Qy 1432 GTGAGGACCGCTCACCAAGAGATGAGGAGAGCATCGAGGAGCTGAGGACCGGCATCG 1491  
Db 1406 GAGATGA---TATCAACAAAGAAACACAGAGCGGTTGATGACTTAAGCCGAGGTATCG 1462  
Qy 1492 TTAACAAGCTCTCTCATGCGCGCTGACGACCTGAGGTGCGAGCGGACGACAGCCGCA 1551

Db 1463 TGAACAGATTCTTACATGTTCCATGTCAGCATTTTGAGATGTGACGGAAGTGCACAGTAGAA 1522  
Qy 1552 CCCTTGAAGAGACGCTCGAGAAACATGACGCGCTTCAACCGGATGTTTTCAGCCTCGACATGG 1611  
Db 1523 CGCTGAGGAGACCGCTTTGAGAAATATGATGCTTTTGAACAGAAATGTCACGGTC- 1576  
Qy 1612 AGAAGGCGATCATCGACGAGAAAGATCAAGGCCAAGGTGGAGAACACAAAACCTGA 1667  
Db 1577 AGAAGACATTTTGGAGGAAAGCTTAAGGCAATGGCGGAACAACAACAGAAAGTAA 1632

Search completed: December 15, 2003, 14:57:28  
Job time : 7206 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 12:46:12 ; Search time 4189 Seconds  
(without alignments)  
11163.001 Million cell updates/sec

Title: US-10-018-902-3  
Perfect score: 1924  
Sequence: 1 ccacgcgcgcgcacatcaataa.....aaaaaaaaaaaaaaaaaaag 1924

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1923	99.9	1923	11	AY106615
c 2	829.4	43.1	831	10	BG319950
3	799.8	41.6	978	13	BQ619299
c 4	755	39.2	800	29	BZ748339

5	752	39.1	752	29	BZ748338
6	685.2	35.6	858	14	CB622460
7	673.8	35.0	854	29	CC424852
8	649.2	33.7	790	14	CB650182
9	632.6	32.9	799	14	CB672611
c 10	622.4	32.3	855	14	CB652061
11	621.4	32.3	812	14	CB621190
c 12	609.6	31.7	845	14	CB684065
13	606	31.5	751	14	CB621670
c 14	601.6	31.3	608	9	AW520155
15	584.2	30.4	733	14	CB685474
16	583.2	30.3	718	9	AU056848
17	583.2	30.3	749	13	BQ483617
c 18	570.6	29.7	786	14	CB670512
19	570.6	29.7	820	14	CB650183
c 20	563.2	29.3	781	14	CB623934
c 21	562	29.2	766	14	CB621191
22	555.8	28.9	559	14	CA452561
23	552.6	28.7	559	14	CA452567
c 24	550.6	28.6	813	14	CB669161
25	526	27.3	582	14	CD206259
26	518.8	27.0	679	13	BQ839101
c 27	508.8	26.4	512	12	B1993101
c 28	507.8	26.4	512	13	BQ539289
c 29	502.8	26.1	703	14	CB648319
c 30	489.6	25.4	543	9	AW671698
31	484	25.2	738	10	BF262195
c 32	481	25.0	719	12	BM084293
33	477.8	24.8	636	9	AV940881
34	476.4	24.8	600	12	B225193
35	471.6	24.5	579	10	BE599448
36	467.4	24.3	621	10	BE366884
c 37	467	24.3	478	12	BG842485
38	464.2	24.1	629	12	BJ452415
c 39	464	24.1	612	10	BE515566
c 40	463.4	24.1	621	14	CB859628
41	463.2	24.1	578	12	BJ214307
42	462.6	24.0	602	9	AV909911
43	461.8	24.0	645	12	B1952816
44	459.6	23.9	579	14	CD206237
45	457	23.8	580	10	BE594855

ALIGNMENTS

RESULT 1	AY106615	AY106615	1923 bp	mRNA	linear	HTC 16-OCT-2002									
LOCUS	Zea mays	PCO073362	mRNA sequence.												
DEFINITION	Zea mays	PCO073362													
ACCESSION	AY106615														
VERSION	AY106615.1	GI:21209693													
KEYWORDS	HTC.														
SOURCE	Zea mays														
ORGANISM	Zea mays														
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.														
AUTHORS	Hayney, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.														
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes														
JOURNAL	Unpublished (2002)														
REFERENCE	2 (bases 1 to 1923)														
AUTHORS	Coe, E.H.														
TITLE	Direct Submission														
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA														
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizegap.org; ZmDB, www.zmdb.iastate.edu; TTGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the														

BZ748338	PUDCK29TB
CB622460	OS11EA09D
CC424852	PURLI39TD
CB650182	OSUNEB14H
CB672611	OSUNEB06L
CB652061	OSJNEC01H
CB621190	OS11EA06O
CB684065	OSJNEF13D
CB621670	OS11EA07L
AW520155	660061E06
CB685474	OSJNEF15O
AU056848	AU056848
BQ483617	WHE3510_F
CB670512	OSJNEE03J
CB650183	OSUNEB14H
CB623934	OS11EA12A
CB621191	OS11EA06O
CA452561	KrIN-2_F0
CA452567	KrIN-2_F0
CB669161	OSJNEE01H
CD206259	HS1_21_D0
BQ839101	WHE3589_D
B1993101	1020072H0
BQ539289	MEST614-D
CB648319	OSJNEB11K
AW671698	LG1_349_B
BF262195	HV_CBA000
BM084293	PPAP_05_C
AV940881	AV940881
B225193	BJ225193
BE599448	P11_87_F0
BE366884	P11_41_F0
BG842485	MEST31-D1
BJ452415	BJ452415
BE515566	WHR0614_B
CB859628	HI11A24W
BJ214307	BJ214307
AV909911	AV909911
B1952816	HVSWEM000
CD206237	HS1_21_D0
BE594855	P11_46_F0

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES

source

Location/Qualifiers  
1..1923  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="MaizeDB:634031"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/DuPont Consensus Library"  
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 504 a 465 c 550 g 404 t  
ORIGIN

Query Match 99.9%; Score 1923; DB 11; Length 1923;  
Best Local Similarity 100.0%; Pred. No. 6.7e-244;  
Matches 1923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACGCGTCCGATCAATAAGAGGAGCTTGGAGTTGCCAAGGCTCCGATTCGCTA 60  
DB 1 CCACGCGTCCGATCAATAAGAGGAGCTTGGAGTTGCCAAGGCTCCGATTCGCTA 60  
QY 61 ATGCGACGATATATGCGACACGAGCTGAGCGACCAACCGCGCGCGAGCGGCGACCA 120  
DB 61 ATGCGACGATATATGCGACACGAGCTGAGCGACCAACCGCGCGCGAGCGGCGACCA 120  
QY 121 CGCCCAAGCGCGGGGTGCTGCTGGCGCTCTGCCAGAGGGTGGCGGGCGGCGGACGC 180  
DB 121 CGCCCAAGCGCGGGGTGCTGCTGGCGCTCTGCCAGAGGGTGGCGGGCGGCGGACGC 180  
QY 181 GCGCTCCGGGTGCTGCGGCTGCGACCGCGCGCGGTGGAGGCCCGAGCGCGCGGTGG 240  
DB 181 GCGCTCCGGGTGCTGCGGCTGCGACCGCGCGGTGGAGGCCCGAGCGCGCGGTGG 240  
QY 241 CAAGGCTCCAGCGCTCGCGCTCGAGCGATTCAGAGATTCGCGCGACCGGTACATGA 300  
DB 241 CAAGGCTCCAGCGCTCGCGCTCGAGCGATTCAGAGATTCGCGCGACCGGTACATGA 300  
QY 301 AGGAAGAGTACCATAGCTGTATAGGCTCTGATAGGCTTACAGAGTACACAGCAGCAGTGGAGATGC 360  
DB 301 AGGAAGAGTACCATAGCTGTATAGGCTCTGATAGGCTTACAGAGTACACAGCAGCAGTGGAGATGC 360  
QY 361 GTGAAAACTTGCTGTCAGAGGAACTGTGGCCCGCTGCTATTCAAGAACTCACTAGCC 420  
DB 361 GTGAAAACTTGCTGTCAGAGGAACTGTGGCCCGCTGCTATTCAAGAACTCACTAGCC 420  
QY 421 TGAACCATATTGAAGAGCTGCTGTTAGTACTGCTGTAATAGAAATTTATGTGG 480  
DB 421 TGAACCATATTGAAGAGCTGCTGTTAGTACTGCTGTAATAGAAATTTATGTGG 480  
QY 481 TGGCGCTATCATGGAACCGTGTATCAGAGAAGTAGTGCATGGAATGTCGAAGAAAGTG 540  
DB 481 TGGCGCTATCATGGAACCGTGTATCAGAGAAGTAGTGCATGGAATGTCGAAGAAAGTG 540  
QY 541 GTATTCCCGCTTCGAGCTTAGGAGCACCTGTTTCATCTTCGGAAGCAGTGTGCCAC 600  
DB 541 GTATTCCCGCTTCGAGCTTAGGAGCACCTGTTTCATCTTCGGAAGCAGTGTGCCAC 600  
QY 601 GCCATCTGTTGAGGTGTCAGCTGGCTTGTGATCTCTTTGGTTCTCGGTGAAGCAAAATCC 660  
DB 601 GCCATCTGTTGAGGTGTCAGCTGGCTTGTGATCTCTTTGGTTCTCGGTGAAGCAAAATCC 660  
QY 661 TTGCTCAGGTTAAACAAGTGTGAGGAGTGACAGAACAGTGGAGGCTTGGGAAGAAC 720  
DB 661 TTGCTCAGGTTAAACAAGTGTGAGGAGTGACAGAACAGTGGAGGCTTGGGAAGAAC 720  
QY 721 TCGATAGGATGTTCAAGGATGCAATCACTGCTGGAAAGCGTGTCCGCGAGGAGACCAACA 780

DB 721 TCGATAGGATGTTCAAGGATGCAATCACTGCTGGAAAGCGTGTCCGCGAGGAGACCAACA 780  
QY 781 TATCATCTGGTGTCTTCTGTAGTTTCTGTCAGCGGGTGTGAACCTGGCCCTGTATGAAGCTTC 840  
DB 781 TATCATCTGGTGTCTTCTGTAGTTTCTGTCAGTTTCTGTCAGCGGGTGTGAACCTGGCCCTGTATGAAGCTTC 840  
QY 841 CGAAGTCTGAAAGCACTGTCACTAGGATGCTTCTGATTTGGTGTGTTTAAATGGGAAAGC 900  
DB 841 CGAAGTCTGAAAGCACTGTCACTAGGATGCTTCTGATTTGGTGTGTTTAAATGGGAAAGC 900  
QY 901 TAGTGATCAACATCTGTTTCCCAAGGATGCAAGAGGTTGTTGGTGTGAACCGCTCCG 960  
DB 901 TAGTGATCAACATCTGTTTCCCAAGGATGCAAGAGGTTGTTGGTGTGAACCGCTCCG 960  
QY 961 TGGAAAGGTTGGATGCTATTCTGTCAGGAGATGAAAGATATAGAGATCGTGTACAGGCCTC 1020  
DB 961 TGGAAAGGTTGGATGCTATTCTGTCAGGAGATGAAAGATATAGAGATCGTGTACAGGCCTC 1020  
QY 1021 TCTCAGACATGATCAAGCTGCTCTGAACTGATGTGTTTCAACAGCACCGCATCTG 1080  
DB 1021 TCTCAGACATGATCAAGCTGCTCTGAACTGATGTGTTTCAACAGCACCGCATCTG 1080  
QY 1081 AAACTTCAATTTTCCCAAGGATGCAAGAGGCTCTCCCTCTGTCTCTGATATATGG 1140  
DB 1081 AAACTTCAATTTTCCCAAGGATGCAAGAGGCTCTCCCTCTGTCTCTGATATATGG 1140  
QY 1141 GAGGTGTTGGCTGCTGTTGTCACATATCTGTGTCAGGATGTGTCAGGCGATGTGTCGTG 1200  
DB 1141 GAGGTGTTGGCTGCTGTTGTCACATATCTGTGTCAGGATGTGTCAGGCGATGTGTCGTG 1200  
QY 1201 AAGTTGCGCTGTCACGAGTGTACAACTGTCGACGCTTGGAAAGAGTGTGGAAGCAACA 1260  
DB 1201 AAGTTGCGCTGTCACGAGTGTACAACTGTCGACGCTTGGAAAGAGTGTGGAAGCAACA 1260  
QY 1261 AGGAGGACCGCTCAGGAAAGCAATGAGGCGGAGCAATCATCAACGAAAGCTGAGGCT 1320  
DB 1261 AGGAGGACCGCTCAGGAAAGCAATGAGGCGGAGCAATCATCAACGAAAGCTGAGGCT 1320  
QY 1321 GGTTCGAGGCAATGAGGAGGCTCGCTGAGGACCGTTCGACCATCAAGAAAGCTGAGGCT 1380  
DB 1321 GGTTCGAGGCAATGAGGAGGCTCGCTGAGGACCGTTCGACCATCAAGAAAGCTGAGGCT 1380  
QY 1381 ACGCGGACAGGATCAGGCGCTCGAGCTCGAGAAAGTGTGTCAGAAAGTGTGAGGAGCG 1440  
DB 1381 ACGCGGACAGGATCAGGCGCTCGAGCTCGAGAAAGTGTGTCAGAAAGTGTGAGGAGCG 1440  
QY 1441 CCCTCACAAGAGATGAGGAGGCTCGAGGAGGCTCGAGGAGGCTCGGATCGTTAAACAGC 1500  
DB 1441 CCCTCACAAGAGATGAGGAGGCTCGAGGAGGCTCGAGGAGGCTCGGATCGTTAAACAGC 1500  
QY 1501 TCCTTCCATGCGCCCTGCGACCTGAGGTGCGAGCGGAGCGAGCGGCGACCCCTTGAAGC 1560  
DB 1501 TCCTTCCATGCGCCCTGCGACCTGAGGTGCGAGCGGAGCGAGCGGCGACCCCTTGAAGC 1560  
QY 1561 AGAGCTCGAGAACATGCAAGCTTCAACCGGATGTTTTCAGCTCGAATGGAAGAGCGA 1620  
DB 1561 AGAGCTCGAGAACATGCAAGCTTCAACCGGATGTTTTCAGCTCGAATGGAAGAGCGA 1620  
QY 1621 TCATCGAGCAGAGATCAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680  
DB 1621 TCATCGAGCAGAGATCAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680  
QY 1681 TTTTCTTACCACTTAT 1740  
DB 1681 TTTTCTTACCACTTAT 1740  
QY 1741 TTTTCTTACCACTTAT 1800  
DB 1741 TTTTCTTACCACTTAT 1800  
QY 1801 GCCATGCGAGGTTGTCATTAACAGGATGTCCTCCCAATTCCTCCCAAGGAGGAGGAGGAG 1860



Db 854 CCCGCAATTCGAAACTCTCATTTGTTCCAAAGAACACCCCGGAGGCACCTTCCC 910

RESULT 4  
 LOCUS BZ748339/c  
 DEFINITION BZ748339 800 bp DNA linear GSS 03-MAR-2003  
 genomic survey sequence.  
 ACCESSION BZ748339  
 VERSION BZ748339  
 KEYWORDS BZ748339.1 GI:28728049  
 SOURCE GSS.  
 ORGANISM Zea mays  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 800)  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick  
 A., Frazer,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.  
 TITLE Maize Genomics Consortium  
 COMMENT Unpublished  
 Other GSSs: PUDCK29TB  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TP  
 Class: sheared ends.  
 Location/Qualifiers  
 1. .800  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMBMTal55F09"  
 /clone\_lib="ZM 0.6 1.0 KB"  
 /notes="Vector: pCRI-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
 Cor selected genomic DNA library"  
 BASE COUNT 174 a 214 c 200 g  
 ORIGIN

Query Match 39.2%; Score 755; DB 29; Length 800;  
 Best Local Similarity 99.1%; Pred. No. 3.5e-90;  
 Matches 780; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

Qy 1114 CACTCCCCCTGCTCTGATACTATGGAGGTGTTCCCTCTTTTTCGACATATCTGTCC 1173  
 Db 799 CTCCTCCCCCTGTTCTCTGATACTATGGAGGTGTTCCCTCTTTTTCGACATATCTGTCC 740  
 Qy 1174 CCA-GGAATGTAG-CGCATGTGTGTCTGAAGTTGGCGCTGCACGAGTGTACAATGTCA 1231  
 Db 739 CCAGGGAATGTCAAGCGCATGTGTGTCTGAAGTTGGCGCTGCACGAGTGTACAATGTCA 680  
 Qy 1232 CGACTTTGAAAGAGGTGCTGGAAGCCAAACAGGAGGACCGGCTCAGGAAGCAATGGAGGC 1291  
 Db 679 CGACTTTGAAAGAGGTGCTGGAAGCCAAACAGGAGGACCGGCTCAGGAAGCAATGGAGGC 620  
 Qy 1292 GCAGACAATCATCACCAGAAAGAACTGAGACGGTTTCGAGGCATGAGAGGAATCTCGCTGGAGAC 1351  
 Db 619 GCAGACAATCATCACCAGAAAGAACTGAGACGGTTTCGAGGCATGAGAGGAATCTCGCTGGAGAC 560  
 Qy 1352 CGTTCGGACCATCAAGAACTGAGGTCTGATCGCGGACGAGTATCGGCTCGGAGTCA 1411  
 Db 559 CGTTCGGACCATCAAGAACTGAGGTCTGATCGCGGACGAGTATCGGCTCGGAGTCA 500  
 Qy 1412 GAAGTCCCTCGCAAAAGTAGGTGAGACGCGCCCTCCACCAAGAAAGATGAGAGAGCCATCGA 1471  
 Db 499 GAAGTCCCTCGCAAAAGTAGGTGAGACGCGCCCTCCACCAAGAAAGATGAGAGAGCCATCGA 440  
 Qy 1472 GAGGCTGAGACGCGGCATCGTTAAACAGCTCTCTCATGCCCGCTCGAGACACTGAGGTG 1531

Db 439 GGAGCTGAGCACCGGCATCGTTAAACAAGCTCTCCATGGCCCGCTGCAGCACCTGAGGTG 380

QY 1532 CGAGCGGACGACAGCGCCACCCCTTGACGAGACGCTCGAGAACATGACGCGCCCTCAACCG 1591

Db 379 CGAGCGGACGACAGCGCCACCCCTTGACGAGACGCTCGAGAACATGACGCGCCCTCAACCG 320

QY 1592 GATGTTTACGCTCGACATCGAGAGAGGCGCATCATCGAGCAGAGATCAAGGCGCAAGGTGGA 1651

Db 319 GATGTTTACGCTCGACATCGAGAGAGGCGCATCATCGAGCAGAGATCAAGGCGCAAGGTGGA 260

QY 1652 GAAGACACAAACTGAGCGCCAGGAGCAATTTTCTACCCACATATCTATATATATAGC 1711

Db 259 GAAGACACAAACTGAGCGCCAGGAGCAATTTTCTACCCACATATCTATATATATAGC 200

QY 1712 GTCTCCCAATCTCAATTCATTTTATCTTTTCACTCAGTCAGCGCCCTTCCCTGCTCACT 1771

Db 199 GTCTCCCAATCTCAATTCATTTTATCTTTTCACTCAGTCAGCGCCCTTCCCTGCTCACT 140

QY 1772 GTGATCGTTAACTGTGCTGTGAATTAGAGCCATGAGCGCATGTTGTCAATAACAGCAAT 1831

Db 139 GTGATCGTTAACTGTGCTGTGAATTAGAGCCATGAGCGCATGTTGTCAATAACAGCAAT 80

QY 1832 GTGTCCCAATTTCCCAAGAGAAAGACTATATTTATATGTCATTTATTTGGAGCAATAGT 1891

Db 79 GTGTCCCAATTTCCCAAGAGAAAGACTATATTTATATGTCATTTATTTGGAGCAATAGT 20

QY 1892 TTACTTA 1898

Db 19 TTACTTA 13

RESULT 5  
B2748338  
LOCUS  
DEFINITION PUDCK29TB\_ZM\_0.5\_1.0\_KB\_Zea\_mays genomic clone ZMMB1a155F09,  
genomic survey sequence.

ACCESSION B2748338  
VERSION B2748338.1 GI:28728048  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick  
A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.  
TITLE Maize Genomics Consortium  
JOURNAL Unpublished  
COMMENT Other GSSs: PUDCK29TD  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: sheared ends.

FEATURES  
source  
1..752  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMB1a155F09"  
/notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
Cor selected genomic DNA library"  
BASE COUNT 198 a 183 c 230 g 141 t

Query Match 39.1%; Score 752; DB 29; Length 752;  
Best Local Similarity 100.0%; Pred. No. 8.9e-90;

Matches 752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 879 GGTCTCGTAAATATGGGAAGCTAGTATCAAAATCTGTTGTCCTCAAGGATGCAAG 938

Db 1 GGTCTCGTAAATATGGGAAGCTAGTATCAAAATCTGTTGTCCTCAAGGATGCAAG 60

QY 939 GTTCTGTGGTGAACCCCTCGTGGAAAGGATGATCTATTTCTGTGAGGAGATGAAGAT 998

Db 61 GTTCTGTGGTGAACCCCTCGTGGAAAGGATGATCTATTTCTGTGAGGAGATGAAGAT 120

QY 999 ATAGAGATCGGTATACGGCTCTCTCAGACATGATCAAGCTGTCTGTGAAGCTGAT 1058

Db 121 ATAGAGATCGGTATACGGCTCTCTCAGACATGATCAAGCTGTCTGTGAAGCTGAT 180

QY 1059 GTGTTTCAACGACCGCATCTGAAATCTTATTTTCCCAAGAACACGAGAGCACTC 1118

Db 181 GTGTTTCAACGACCGCATCTGAAATCTTATTTTCCCAAGAACACGAGAGCACTC 240

QY 1119 CCCCTGTCTGTATCTATATGGGAGGTGTTGCGCTGTTTTCGACATATCTGTCCCCAGG 1178

Db 241 CCCCTGTCTGTATCTATATGGGAGGTGTTGCGCTGTTTTCGACATATCTGTCCCCAGG 300

QY 1179 AATGTACGCGCATGTGTCTGAAGTTGGCGCTGCACAGTGTACAATGTGACGACTTG 1238

Db 301 AATGTACGCGCATGTGTCTGAAGTTGGCGCTGCACAGTGTACAATGTGACGACTTG 360

QY 1239 AAAGAGTGTGGAAAGCAACAGGAGGACCGGCTCAGGAAGCAATGGAGGCGCAGACA 1298

Db 361 AAAGAGTGTGGAAAGCAACAGGAGGACCGGCTCAGGAAGCAATGGAGGCGCAGACA 420

QY 1299 ATCATCACCAAGAACTGAGACGCTTCGAGGCAATGGAGGAGCTCGTGGAGACGTTCCG 1358

Db 421 ATCATCACCAAGAACTGAGACGCTTCGAGGCAATGGAGGAGCTCGTGGAGACGTTCCG 480

QY 1359 ACCATCAAGAACTGAGTGTGTAGCGGACAGGATCAGGGCTTCGAGGCTCGAGAAGTGC 1418

Db 481 ACCATCAAGAACTGAGTGTGTAGCGGACAGGATCAGGGCTTCGAGGCTCGAGAAGTGC 540

QY 1419 CTCGAGAAAGTGTGAGGAGCGCTTCACCAAGAAATGAGGAGGAGCCATCGAGAGCTG 1478

Db 541 CTCGAGAAAGTGTGAGGAGCGCTTCACCAAGAAATGAGGAGGAGCCATCGAGAGCTG 600

QY 1479 AGCACCGGATCGTTTAAACAGCTCTCCATGGCCCGCTGCAGACCTGAGGTGGAGCGC 1538

Db 601 AGCACCGGATCGTTTAAACAGCTCTCCATGGCCCGCTGCAGACCTGAGGTGGAGCGC 660

QY 1539 AGCGACAGCGCACCCCTTTCAGGAGAGCTTCGAGAACATGCAAGCCCTCAACCGGATGTC 1598

Db 661 AGCGACAGCGCACCCCTTTCAGGAGAGCTTCGAGAACATGCAAGCCCTCAACCGGATGTC 720

QY 1599 AGCTTCGACATGAGGAAGGCGCATCTCGAGCA 1630

Db 721 AGCTTCGACATGAGGAAGGCGCATCTCGAGCA 752

RESULT 6  
CB622460  
LOCUS  
DEFINITION OSIIIEa09D05.f OSIIIEa Oryza sativa (indica cultivar-group) cDNA  
clone OSIIIEa09D05 5', mRNA sequence.

ACCESSION CB622460  
VERSION CB622460.1 GI:29617448  
KEYWORDS EST.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoidae; Oryzae; Oryza.  
REFERENCE 1 (bases 1 to 858)  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea

JOURNAL  
COMMENT

Unpublished  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 09 row: D column: 05  
Seq primer: gta aaa cga cgg cca gtcg.

## FEATURES

source  
1..858  
Location/Qualifiers  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="IR36"  
/db\_xref="taxon:39946"  
/clone="OS11Ea09D05"  
/tissue\_type="Leaf"  
/dev\_stages="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OS11Ea"  
/note="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2:  
XhoI; Lesion Mimic SPL11"  
XhoI: 216 a 171 c 245 g 226 t

BASE COUNT 216 a 171 c 245 g 226 t  
ORIGIN

Query Match 35.6%; Score 685.2; DB 14; Length 858;  
Best Local Similarity 87.4%; Pred. No. 5e-81;  
Matches 750; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 359 CCGTGAATAAATCTGCTTTCAGAGAACTGTGGCCCGTCTATTCAGAACTCCTAG 418  
Db 1 CGGTGAGAAATCTGCTTTCAGAGAACTGTGGCCCGTCTATTCAGAACTCCTAG 60  
QY 419 CTTGAACCATATGAGAGGCTGCTTTCAGAGAACTGTGGCCCGTCTATTCAGAACTCCTAG 478  
Db 61 TCTGATCATATGAGAGGCTGCTTTCAGAGAACTGTGGCCCGTCTATTCAGAACTCCTAG 120  
QY 479 GGTGGCGTATCATGGAACCGTGTATCAGAGAACTGTGGCCCGTCTATTCAGAACTCCTAG 538  
Db 121 GGTAGCTTATCGTGAACCGTGTATCAGAGAACTGTGGCCCGTCTATTCAGAACTCCTAG 180  
QY 539 TGGTATTCCTCGCTTCAGAGCTAGGAGCACTGTTTCATCTTGGAGAGCTGTATGCCAC 598  
Db 181 TGGAAATCCCTGCTCTGAGCTCAGGAGCATCTATTCATGTTGCGTGACAGTGTGCCAC 240  
QY 599 ACGCCATCTGTTGAGGTGTGAGCTGGCTTTCAGTCTTTCAGTCTTTCAGTGAAGCAAAAT 658  
Db 241 ACGCCATCTGTTGAGGTGTGAGCTGGCTTTCAGTCTTTCAGTGAAGGCGCAAT 300  
QY 659 CTTGCTCAGGTTAAACAAGTTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 718  
Db 301 CTTGCTCAGTAAACAAGTTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 360  
QY 719 CATCGATAGGATGTTCAGGATGCAATCAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 778  
Db 361 CATCGATAGGATGTTCAGGATGCAATCAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 420  
QY 779 CATATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 838  
Db 421 CATATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
QY 839 TCCGAAGTCTGAACACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 898  
Db 481 TCCAAAGTCTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
QY 899 GCTAGTGTATCAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 958  
Db 541 GTTGGTGTAAACATTTGTTCCAAAGGATGCAAGAAAGTGTGTTGTTGTTGTTGTTGTTG 600

QY 959 CGTGAAGAGGTGATCTATTCAGAGAGATGAAGAGATATAGAGATCTGTTGACAGGCC 1018  
Db 601 AGTGAAGAGGTGATCTATTCAGAGAGATGAAGAGATATAGAGATCTGTTGACAGGCC 660  
QY 1019 TCTCTCAGACATGATCAAGCTGCTGCTGAGAGCTGATGCTGTTTCCAGCAGCAGCATC 1078  
Db 661 TCTTACAGAGATGATGAAGCCGCTGCCAGAGCTGATGCTGTTTCAAGCAGCAGCATC 720  
QY 1079 TGAACCTTCATTTTCGCAAGAACACGACAGAGCACTCCCTCTGCTCTGATCTAT 1138  
Db 721 CGAAACCCCATTTTCACAAAGAGAGCAGCAGAGCGCTTCCCGCTATTTCTGATGCTAT 780  
QY 1139 GGGAGGTGTTGCTGCTGTTTTCGACATATCTGTTCCAGAGAAATGTACAGCGATGTGTC 1198  
Db 781 GGGTGGTGTTCGACTCTTTTTCGACATATCTGTTCCAGAGAAATGTACAGCGCTGTGTC 840  
QY 1199 TGAAGTTGGCGCTGCACG 1216  
Db 841 TGAAGTTGGCATCGCG 858

RESULT 7  
CC424852 854 bp DNA linear GSS 19-MAY-2003  
LOCUS PUHLI39TD ZM 0.6\_1.0 KB Zea mays genomic clone ZMMB7a477H05,  
DEFINITION genomic survey sequence.  
ACCESSION CC424852  
VERSION CC424852.1 GI:30904942  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE 1 (bases 1 to 854)  
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick  
A., Frazer,C.M., Yuan,Y., San Miguel,P., Ma,J. and Benmetzen,J.  
TITLE Maize Genomics Consortium  
COMMENT Other GSSs: PUHLI39TB  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: Tg  
Class: sheared ends.

## FEATURES

## source

1..854  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMB7a477H05"  
/clone\_lib="ZM\_0.6\_1.0 KB"  
/notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
Cot selected genomic DNA library"  
BASE COUNT 226 a 208 c 245 g 175 t  
ORIGIN

Query Match 35.0%; Score 673.8; DB 29; Length 854;  
Best Local Similarity 88.4%; Pred. No. 1.6e-79;  
Matches 754; Conservative 0; Mismatches 97; Indels 2; Gaps 2;

QY 750 GCTGGAAGCGTGTCCGACGAGAGCAACATATCATCTGCTGCTGCTGCTGCTGCTGCTG 809  
Db 2 GCGGGAAGCGTGTGCGCTCCGAGACTAACATATCATCTGCTGCTGCTGCTGCTGCTGCTG 61  
QY 810 GCGGCGGTTGAACCTGGCCCTGTGATGAAGCTTCCGAAGTCTGAAGCACTGTGCTGCTG 869  
Db 62 GCGGCGGTTGAATTTGGCCCTGTGATGAAGCTTCCGAAGTCTGAAGCACTGTGCTGCTGCTG 121

```
QY 870 CTTCTGATGGTGTGTAATAATGGAAAGCTAGTGATCAAAACATCTGTTGCCAAAGCA 929
Db |||
QY 122 CTTTGTGTTGGTGGCGGCAAAATGGAAACCTAGTGATCAAAACCTGATTCCTCAAGCA 181
Db |||
QY 930 TCAAGAGAGTGTGTTGGTGAACCGCTCCGTGGAAAGGTTGATGCTATTCTGTGAGGAG 989
Db |||
QY 182 TGCAAAAAGGTGGTGTGTTGTTCAACCGTTTCAAGCAAAAGGTTGATGCTATCCGCGAGGAG 241
QY 990 ATGAAGATATAGAGATCGTGTACAGAGCTCTCTCAGACATGATATCAAGCTGCTCTGAA 1049
Db |||
QY 242 ATTAAGCTATCGAGATGTGTACAGGCTCTCTCGAGATGATGAAGCTGCTCTGAA 301
QY 1050 GCTGATGTGTTTCCACAGCAGCCGATCTGAAACTTTCATTTGTCGAAAGAACACGCA 1109
Db |||
QY 302 GCTGACGTGTTTCAAGAGCAGCCGATCTGAACCCCATTTTCAAAAGAGACGCA 361
QY 1110 GAGGCACTCCCTCTCTGATACTATGAGAGTGTTCGCTGTTTGTGACATATCT 1169
Db |||
QY 362 GATGACCTTCCACTGTTTCTAATGCCATGGCGGTTCGCGCTCTTTGTGACATATCT 421
QY 1170 GTCCCCAGGAATGTCAAGCGCATGTGTCTGAAGTTGGCTGTCAGAGTGATACAAATCTC 1229
Db |||
QY 422 GTCCCCAGGAATGTCAAGCGCTGTGTCTGAAATTTGCTCCGCGAGTGATCAATGTT 481
QY 1230 GAC3ACTTTGAAAGAGTGTGGAAGCAACAAAGAGAGACCGCTCAGGAAAGCAATCGAG 1289
Db |||
QY 482 GATGACCTGAAAGAGTGTGGAAGCAACAAAGAGAGACCGCTCAGGAAAGCAATCGAG 541
QY 1290 GCGGAGACATCATCAGCAAGAACTGAGACGGTTTCGAGGAGTGAAGGAGCTCGCTGGAG 1349
Db |||
QY 542 GCGGAGACATCATCAGCAAGAGTGAACGGTTTGAAGGATGGCGGACTCGCTGGAG 601
QY 1350 ACCGTTCCGACCATCAAGAGCTGAGTGTGTAAGTGGCTGACGAGACGATCAGGCGCTCGAGCTC 1409
Db |||
QY 602 ACCGTTCCACCATCAAGAGCTGAGTGTGTAAGTGGCTGACGAGACGATCAGGCGCTCGAGCTC 661
QY 1410 GAGAGTCCCTGCAGAA - AGTAGTGGAGAGCGCTCACCAGAGAGATGAGGAGAGCCAT 1468
Db |||
QY 662 GAGAGTCCCTGCAGAAAGTGGGATGAGCTTCTACCAAGAGATGAGGAGAGCCAT 721
QY 1469 CGAGAGCTGAGCAGCGGATGTTTAAAGTCTCTCATGCGCTGCGAGCAGCTGAG 1528
Db |||
QY 722 CGAGAGCTGAGCAGCGGATGTTTAAAGTCTCTCATGCGCTGCGAGCAGCTGAG 781
QY 1529 GTGGAGCGGAGCAGCGGATGTTTAAAGTCTCTCATGCGCTGCGAGCAGCTGAG 1587
Db |||
QY 782 GTGGAGCGGATGAGCAGCGGATGTTTAAAGTCTCTCATGCGCTGCGAGCAGCTGAG 1641
QY 1588 ACCGATGTTTCAAG 1600
Db |||
QY 842 ACCGATGTTTCAAG 854
```

```
RESULT 8
LOCUS CB650182
DEFINITION OSJNEB14H23.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEB14H23 5', mRNA sequence.
ACCESSION CB650182
VERSION CB650182.1
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 790)
AUTHORS Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
TITLE Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL Unpublished
```

```
CONTACT: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR PRIMERS
FORWARD: gta aaa cga cga cga gta gta
BACKWARD: gga aac agc tat gac cat g
Plate: 14 row: H column: 23
Seq primer: gta aaa cga cga cga gta gta
Location/Qualifiers
source
1..790
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB14H23"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNBb"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)"
BASE COUNT 208 a 154 c 231 g 197 t
ORIGIN
Query Match 33.7%; Score 649.2; DB 14; Length 790;
Best Local Similarity 88.9%; Pred. No. 2.8e-76;
Matches 702; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 237 GTGCAAGAGGTCGAGCGCTCGCGCCCTCCAGCAGTTCAGAGATATCCGCGACCGGTAC 296
Db 1 GTGCAAGAGGTCGAGCGCTCGCGCCCTCCAGCAGTTCAGAGATATCCGCGACCGGTAC 60
QY 297 ATGAAGAAAGAGTACATAGCTGTGATAGGCTCAGTGTACACAGCAGCAGTGTGAG 356
Db 61 ATGAAGAAAGAGTACATAGCTGTGATAGGCTCAGTGTACACAGCAGCAGTGTGAG 120
QY 357 ATGCGTGAAGAACTTGTGTTGAGAGGACTGTGGCCCGTGTATTCAGAGACTCACT 416
Db 121 ATGCGTGAAGAACTTGTGTTGAGAGGACTGTGGCCCGTGTATTCAGAGACTCACT 180
QY 417 AGCCTGAACCATATTGAAGAGGCTGTGTTCTTAGTACCTGTAATAGAAATTTAT 476
Db 181 AGTCTGAATCATATTGAAGAGGCTGTGTTCTTAGTACCTGTAATAGAAATTTAT 240
QY 477 GTGTCGGCTATCATGGAACCGTGTGTATCAGAGAGTGTGAGAGTGTGAGAGAA 536
Db 241 GTGTCGGCTATCATGGAACCGTGTGTATCAGAGAGTGTGAGAGTGTGAGAGAA 300
QY 537 AGTGTATTCGCGCTCCGAGCTTAGGAGCAGCTGTTCATCTTTCGAGAGCAGTGTGCC 596
Db 301 AGTGTATTCGCGCTCCGAGCTTAGGAGCAGCTGTTCATCTTTCGAGAGCAGTGTGCC 360
QY 597 ACAGCCCATCTGTTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 656
Db 361 ACAGCCCATCTGTTTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 420
QY 657 ATCTTGTCTCAGGTTTAAACAAAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 716
Db 421 ATCTTGTCTCAGGTTTAAACAAAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 480
QY 717 AACATGATAGGATGTTCAAGGATGCAATCACTGTGGAAGGCTGTCCGAGCAGCAG 776
Db 481 AACATGATAGGATGTTCAAGGATGCAATCACTGTGGAAGGCTGTCCGAGCAGCAG 540
QY 777 AACATATCATCTGCTGCTGTTTCTGTCAGTTCAGCGCGGTTCAGTTCGCGCTGTGAG 836
Db 541 AACATATCATCTGCTGCTGTTCTCTGTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 600
```

COMMENT

```

Qy 837 CTTCCGAAGTCGAAGCACTGTCACTAGGATGCTTCTGATTGGTGGTAAATGGGA 896
Db 601 CTTCCAAAGTCGAATGCCATCTCTGTAGGATGCTGTTGATTGGTGGCAAGATGGGA 660

Qy 897 AAGCTAGTAGTCAAAACATCTGTTCCCAAGGATGCAAGAGGTTGTTGGTGAACCGC 956
Db 661 AAGTTGGTGGTAAACATTTGATTGCCAAGGATGCAAGAAAGTTGTTGGTGAACCGT 720

Qy 957 TCCGTGGAAAGGTTGGATGCTAATTCGTGAGGAGATGAAAGATATAGAGATCGTGTACAGG 1016
Db 721 TCAGTGGAAAGGTTGGATGCTAATTCGTGAGGAGATGAAAGATGAGATTGTGTACAGG 780

Qy 1017 CCTCTCTCAG 1026
Db 781 CCTCTTACAG 790

RESULT 9
CB672611
LOCUS
DEFINITION CB672611.1 799 bp mRNA linear EST 09-APR-2003
clone OSJNEe06L17.5', mRNA sequence.
ACCESSION CB672611
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 799)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 06 row: L column: 17
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1..799
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEe06L17"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEe"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (70-15)"
BASE COUNT 202 a 156 c 229 g 212 t
ORIGIN

Query Match 32.9%; Score 632.6; DB 14; Length 799;
Best Local Similarity 87.0%; Pred. No. 4.2e-74;
Matches 695; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 444 GTTCTTACTACTGTAATAGATGCAATTTATGCTGGCGCTATCATGGAACCGTGGT 503
Db 1 GTTCTTACTCTGCTTATATGGAATCTATGTGTAGCTTTATCTGGAACCGTGGG 60

```

```

Qy 504 ATCAGAGAGTAGTGGACTGATGTGCAAGAAAAGTGTATTTCCGCTTCCGAGCTTAGG 563
Db 61 ATTAGAGAAGTGGTAGACTGGATGCTCAAAGAAAAGTGGAAATCCCTGCTTCTGAGCTCAGG 120

Qy 564 GAGCACCTGTTTCATCTTCGGAAGCAGTGTATGCCACACGCCATCTGTTTGGAGGTGTCAAGCT 623
Db 121 GAGCATCTATTTCATGTTGCGTGACAGTGTATGCCACACGCCATCTGTTTGGAGGTATCTGCT 180

Qy 624 GGCCTTGATCTCTTTGGTTCTCGGTGAAGGACAAATCTCTTGTCTCAGGTTAAACAAGTTGTG 683
Db 181 GGGCTTGACTCTTTGGTTCTTTGGAGAGAGGGCAATCTCTTGTCTCAAGTTAAACAAGTTGTG 240

Qy 684 AGGAGTGGACAGAACAGTGGAGGCTTGGGAAAAGAACATCGATAGATGATTTCAAGATGCA 743
Db 241 AGAAGTGGGCAAAAACAGTGGAGGCTTGGGAAAAGAACATCGATAGATGATTTCAAGATGCA 300

Qy 744 ATCACTGCTGAAAAGCGTGTCCGAGCGAGACCAACATATCATCTGGTGTCTGTTCTGTCTGTC 803
Db 301 ATCACTGCTGAAAAGCGTGTCCGAGCGAGACTAACATATCATCAGGTGCTGTCTGTCTGTC 360

Qy 804 AGTTACGCGCGGTGAACACTGCCCTGATGAAGCTTCGGAAGTCTGAAGCACTGTCAAGCT 863
Db 361 AGTTACGCTGCAGTTGAATTTGGCTTGTATGAAGCTTCCAAAGTCCGAATGCCATCTGCT 420

Qy 864 AGGATGCTTCTGATTGGTGTCTGTAATAATGGGAAAGCTAGTATCAAAACATCTGGTGTGCC 923
Db 421 AGGATGCTTGTGATTGGTGTCTGCAAGATGGGAAAGTTGGTGTGTTAAACATTTGATTGCC 480

Qy 924 AAAGGATGCAAGAAAGTTGTTGGTGAACCGCTCCGTGGAAAGGTTGGATGCTATTTCGT 983
Db 481 AAGGATGCAAGAAAGTTGTTGGTGAACCGCTTCAGTGGAAAGGTTGGATGCTATCCGCTCCG 540

Qy 984 GAGGAGATGAAAGATATAGAGATCGTGTACAGGCTCTCTCAGACATCATCAAGCTGCT 1043
Db 541 GAAGAGATGAAAGACATTTGAGATTGTGTACAGGCTCTCTTACAGAGATGTATGAAGCCGCT 600

Qy 1044 GCTGAAGCTGATGCTGTTTCCACGACGCCATCTGAACTTCAATTTCTGCAAAAGNA 1103
Db 601 GCCGAAGCTGATGCTGTTTCCACGACGCCATCTGCAAAAGCCCAATTTCTCACAAGGAG 660

Qy 1104 CACGACAGGCACTCCCCCTGCTCTGATATACTATGGAGGCTGTTGCGCTGTTTCTGCAAC 1163
Db 661 CACGACAGGCGCTTCCCGCTATTTCGTATGCTATGAGTGTGCTGCACTCTTTGTCGAC 720

Qy 1164 ATATCTGTCCCAGGAATGTTCAGCGCATGTGTGCTGAAGTTGGCGTGCACGAGTGTAC 1223
Db 721 ATATCCGTCGCCAGAAATGTTCAGCGCTGTGTGCTGAAGTTGGCGCATGCGCGAGTATAC 780

Qy 1224 AATCTCGACGACTTGAAG 1242
Db 781 AACGTGATGACTTGAAG 799

RESULT 10
CB652061/c
LOCUS
DEFINITION CB652061.1 855 bp mRNA linear EST 09-APR-2003
clone OSJNEc01H23.3', mRNA sequence.
ACCESSION CB652061
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 855)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished

```



COMMENT

Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: <http://genome.arizona.edu>  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtcg  
BACKWARD: gga aac agc tat gac cat g  
Plate: 01 row: H column: 23  
Seq primer: gga aac agc tat gac cat g  
Location/Qualifiers  
1. .855  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNEC01H23"  
/tissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSJNEC"  
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"  
BASE COUNT 186 a 231 c 205 g 233 t  
ORIGIN

FEATURES

source  
Query Match 32.3%; Score 622.4; DB 14; Length 855;  
Best Local Similarity 85.6%; Pred. No. 8.9e-73;  
Matches 692; Conservative 0; Mismatches 116; Indels 0; Gaps 0;  
Qy 912 CATCTGTTGCAAGGATGCAAGAGTGTCTGCTGACCGCTCGGTGGAAAGGTTG 971  
Db 855 CATTTGATGCCAAGGATGCAAGAGTGTCTGCTGACCGCTCGGTGGAAAGGTTG 796  
Qy 972 GATGCTATTCTGAGGAGATGAAAGATATAGAGATCGTGTACAGCGCTCTCTCAGACATG 1031  
Db 795 GATGCCATCCGGAAGAGATGAAAGATGAGATGTTGTACAGGCTCTTACAGAGATG 736  
Qy 1032 TATCAAGCTGCTGAAGCTGATGCTGTTTACACGACCGCATCTGAAACTTCATTG 1091  
Db 735 TATGAACCGCTGCGAAGCTGATGCTGTTTACACGACCGCATCCGAAACCCCATTTG 676  
Qy 1092 TTCCAAAGAACACGACGAGGACATCCCGCTGCTCTCATATGAGGAGTTCGC 1151  
Db 675 TTCAAAAGAGACGACGAGGAGGCTTCCCGCTATTCTGATGCTATGGGTGTTTCA 616  
Qy 1152 CTGTTTGTGACATATCTGTCCCGAGGAATGTGAGCGCATGTGTCTGAAGTTGGCGCT 1211  
Db 615 CTCTTTGTGACATATCCGTCCCGAGAAATGTGAGCGCTGTGTCTGAAGTTGGCCAT 556  
Qy 1212 GCAGGAGTGTACATGTGAGGAGTGTGAAAGAGTGTGGAAGCCAAAGGAGGACCGG 1271  
Db 555 GCOCGAGTATACACGCTGATGATCTTGAAGAGAGTGTGGAAGCCAAAGGAGGACCGG 496  
Qy 1272 CTCAGGAAAGCAATGAGGAGGCGACAAATCATCACGAGAACTGAGCGGTTTCAGGCA 1331  
Db 495 CTTAGGAAAGCAATGAGGAGGCGCAAAATCATCACCAAGATGGAACGGTTTCAGGCA 436  
Qy 1332 TGAAGGAGCTCGTGGAGACCGTTTCCGACCATCAAGAAGCTGAGGTGTTACGCGGACAGG 1391  
Db 435 TGAAGGAGCTCGTGGAGACTGTTCCGACTATCAAGAAGCTGAGGTGTTACGCGGACAGG 376  
Qy 1392 ATCAGGCGCTCGAGCTCGAAGAGTGTCTGCAAAAGTAGGTGAGGACGCGCTTCAACAAG 1451  
Db 375 ATCAGGCGCTCGAGCTTGAAAGAGTGTCTGCAAAAGTAGGTGAGGACGCGCTTCAACAAG 316  
Qy 1452 AAGATGAGGAGACCATCGAGGAGCTGAGCAGCGGATCGTTAACAGCTCTCTCCATGCG 1511  
Db 315 AAGATGAGGAGATCCATCGAGGAGCTGAGCAGCGGATCGTTAACAGCTCTCTCCACGGC 256

Qy 1512 CCGCTGCAGCACCTGAGTGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1571  
Db 255 CCATTGCAGCACCTGAGTGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 196  
Qy 1572 AACATGCAAGCGCTCAACCGGATGTTTCAGCTCGACATGGAGAGGAGGAGGAGGAGGAGGAG 1631  
Db 195 AACATGCAAGCGCTCAACCGGATGTTTCAGCTCGACATGGAGAGGAGGAGGAGGAGGAGGAG 136  
Qy 1632 AAGATCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1691  
Db 135 AAGATCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 76  
Qy 1692 CCATTATCTATATATATAGGCTCTCCAA 1719  
Db 75 TCAGCCCGTGTACTATGTATATA 48

RESULT 11

CB621190  
LOCUS CB621190  
DEFINITION Oryza sativa (indica cultivar-group) cDNA  
clone OSIIa06012.5', mRNA sequence.  
CB621190  
ACCESSION CB621190.1 GI:29616178  
KEYWORDS EST.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartidae; Oryzaceae; Oryza.  
1 (bases 1 to 812)  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
Unpublished  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: <http://genome.arizona.edu>  
PCR Primers  
FORWARD: gta aaa cga cgg cca gtcg  
BACKWARD: gga aac agc tat gac cat g  
Plate: 06 row: O column: 12  
Seq primer: gta aaa cga cgg cca gtcg.  
Location/Qualifiers  
1. .812  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="IR36"  
/db\_xref="taxon:39946"  
/clone="OSIIa06012"  
/tissue\_type="Leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSIIa"  
/note="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2:  
XhoI; Lesion Mimic SPL 11"  
BASE COUNT 192 a 171 c 258 g 190 t  
ORIGIN

FEATURES

source  
Query Match 32.3%; Score 621.4; DB 14; Length 812;  
Best Local Similarity 87.4%; Pred. No. 1.2e-72;  
Matches 692; Conservative 0; Mismatches 97; Indels 3; Gaps 1;  
Qy 165 GCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 221  
Db 21 GCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 80

Query Match

Best Local Similarity 87.4%; Pred. No. 1.2e-72;  
Matches 692; Conservative 0; Mismatches 97; Indels 3; Gaps 1;  
Qy 165 GCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 221  
Db 21 GCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 80



```

Db      245  GATGTGACGGCAGCAGCCGCTCGATGAGACGCTGAGAACATGACGCGCTCA 186
QY      1588  ACCGATGTTTACGCTTCGACATGGAGAGCCATCATCGAGCAGAGATCAAGGCCAAGG 1647
Db      185  ACAGGATGTTACGCTTCGACACCGAGAGAGCCATCATGAGCAGAGATCAAGGCCAAGG 126
QY      1648  TGGAGAAGACACAAACTGAGGCCAGGAGCAATTTTCTACCAACCATTTATCTATATATA 1707
Db      125  TGGAGAAGTCCAGAACTGAGATTGAAGAAGAGATTTTTTTTTTTCAGCCCGTGTATCTA 66
QY      1708  TAGCGTCTCCAA 1719
Db      65  CTATGTATACTA 54

```

## RESULT 13

```

CB621670      751 bp mRNA linear EST 08-APR-2003
OSIIeA07L19.f OSIIeA Oryza sativa (indica cultivar-group) cDNA
clone OSIIeA07L19 5', mRNA sequence.
CB621670

```

```

ACCESSION      CB621670
VERSION        CB621670.1 GI:29616658

```

## KEYWORDS

## SOURCE

## ORGANISM

```

Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

```

## REFERENCE

## AUTHORS

```

Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.

```

## TITLE

```

Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea

```

## JOURNAL

## COMMENT

```

Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

```

```

PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 07 row: L column: 19
Seq primer: gta aaa cga cgg cca gtc.

```

## FEATURES

## source

```

1..751
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/clone="OSIIeA07L19"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSIIeA"
/notes="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; Lesion Mimic SPL 11"

```

## BASE COUNT

## ORIGIN

```

196 a 142 c 217 g 196 t
Query Match      31.5%; Score 606; DB 14; Length 751;
Best Local Similarity 88.0%; Pred. No. 1.4e-70;
Matches 660; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```

```

QY      331  TCAGTGTACACACAGCACCAGTGGAGATCGCTGAAAACTTGCTGTGACAGGAACGTGT 390
Db      2    TCAGTGTACACCCCTTACCAGTGGAGATCGCTGAGAACTTGCTGTGACAGGAAC 61
QY      391  GGCCCCGTGCTATTCAAGAACTCACTAGCCTGAAACCATATTGAAGAGGCTGCTGTTCTTA 450

```

```

Db      62  GGGCCCGTGCTATCTCAGAACTCACCAGTCTGAATCATATTGAAGGCGTCTGTTCTTA 121
QY      451  GTACCTGTAAATAGAAATTTATGTGTGGCGCTATCATGGAACCGTGTGTATCAGAG 510
Db      122  GTACCTGCAATAGAAATGGAATCTATGTGTAGTCTTTATCGTGGAAACCGTGGATTAGAG 181
QY      511  AAGTAGTGGCATGGATGTCGAAGAAAGTGGTATTCCCGCTTCCGAGCTTACGGAGCACCC 570
Db      182  AAGTGTAGATCTGGATGTCAAAGAAAGTGGATCCCTGCTTCTGAGCTCAGGAGCATC 241
QY      571  TGTTCATCTTGCAGAGCAGTGTATGCACACGCCATCTGTTTGGAGTGTGAGCTGGCGTTG 630
Db      242  TATTCATGTTGCGTGACAGTGTATGCCACACGCCATCTGTTTGGAGTATCTGCTGGCTTG 301
QY      631  ACTCTTTGGTCTCGGTGAAGGACAAATCTCTTGTCTCAGGTTTAAACAAGTTGTGAGAGTG 690
Db      302  ACTCTTTGGTCTTGGAGAGGCGCAATCTCTTGTCTCAAGTTTAAACAAGTTGTGAGAGTG 361
QY      691  GACAGAACAGTGGAGGCTTGGGAAGAACATCGATAGGATGTTCAAGGATGCAATCAGCTG 750
Db      362  GGCAAAACAGTGGAGGCTTGGGAAGAACATCGATAGGATGTTCAAGGATGCAATCAGCTG 421
QY      751  CTGGAAGCGTGTCCGACGAGAGCAACATATCATCTGGTCTGCTTTCTGTGAGTTCAG 810
Db      422  CTGGAAGCGTGTCCGCTGCGAGACTAACATATCATCAGTGTCTCTCTGTGAGTTCAG 481
QY      811  CGGCGGTTGAATCGGCCCTCGATGAAGCTTCCGAAGTCTGAAGCAGCTGTCAGTAGGATGC 870
Db      482  CTGCAGTTGAATTTGGCTTGTATGAAGCTTCCAAAGTCCGAATGCCCTATCTGCTAGGATGC 541
QY      871  TTCTGATTGGTGGTAAATGGGAAGCTAGTATCAACATCTGCTGTCGAAGGAT 930
Db      542  TGTGATTGGTGGTGAAGATGGGAAGTGGTGGTGTAAACATTTGATTTGCCAAGGGAT 601
QY      931  GCAAGAAGTGTGTTGTGTGAACCGCTCCGCGGAAGGTTGGATGATCTTCGTGAGGAGA 990
Db      602  GCAAGAAGTGTGTTGTGTGAACCGTTCAGTGGAAAGGTTGGATGCCATCCGCGAAGAGA 661
QY      991  TGAAGATATAGAGATCGTGTACAGGCTCTCTCAGACATGTATCAAGCTGCTGCTGAAG 1050
Db      662  TGAAGACATTTGAGATTGTGTACAGGCTCTTTACAGAGATGTATGAAGCCGCTGCCGAAG 721
QY      1051  CTGATGCTGCTGTTCCACCAGCACCCCATCTG 1080
Db      722  CTGATGCTGCTGTTCCACAGACCGCATCCG 751

```

## RESULT 14

## LOCUS

## DEFINITION

```

AW520155      60061E06.xl 660 - Mixed stages of anther and pollen Zea mays cDNA,
mRNA sequence.

```

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

```

AW520155      60061E06.xl 660 - Mixed stages of anther and pollen Zea mays cDNA,
mRNA sequence.
AW520155      60061E06.xl 660 - Mixed stages of anther and pollen Zea mays cDNA,
mRNA sequence.
AW520155.1    GI:7162533
EST.
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 608)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu

```

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

```

FEATURES
  source
    Plate: 660061 row: E column: 06.
    Location/Qualifiers
      1..608
        /organism="Zea mays"
        /mol_type="mRNA"
        /cultivar="Ohio43"
        /db_xref="taxon:4577"
        /tissue_type="whole premeiotic anthers to pollen shed"
        /dev_stage="premeiotic anthers to pollen shed"
        /lab_host="XLOUR"
        /clone_lib="660 - Mixed stages of anther and pollen"
        /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
        Site_2: XhoI; Anther and pollen cDNA library.
        Directionally sequenced with 5' end at the EcoRI site.
        Created by Amie Franklin."
      BASE COUNT      114 a 170 c 158 g 166 t
      ORIGIN
        Query Match      31.3%; Score 601.6; DB 9; Length 608;
        Best Local Similarity 99.3%; Pred. No. 5.9e-70;
        Matches 604; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
      QY 1179 AATGTCAGCGCATGTGTCTGAAGTTGGCGCTGCACGAGTGTACAATGTCAGCACTTG 1238
      Db 608 AATGTCAGCGCATGTGTCTGAAGTTGGCGCTGCACGAGTGTACAATGTCAGCACTTG 549
      QY 1239 AAAGAGGTGGTGAAGCAACCAAGGAGGACCGCTCAGGAAGCAATGAGGCGCAGACA 1298
      Db 548 AAAGAGGTGGTGAAGCAACCAAGGAGGACCGCTCAGGAAGCAATGAGGCGCAGACA 489
      QY 1299 ATCATCACCAGAAACTGAGACGGTTCGAGGATCGAGGGACTCCGCTCGAGACCGCTTCGG 1358
      Db 488 ATCATCACAAGAACTGAGACGGTTCGAGGATCGAGGGACTCCGCTCGAGACCGCTTCGG 429
      QY 1359 ACCATCAAGAGCTGAGGTGTCACGGACAGGATCAGGGCTTCGAGCTCGAGAAGTGC 1418
      Db 428 ACCATCAAGAGCTGAGGTGTCACGGACAGGATCAGGGCTTCGAGCTCGAGAAGTGC 369
      QY 1419 CTGCAAGAAAGTAGGTGAGAGCGCCCTCACCAGAAAGATGAGGAGGCCATCGAGGAGCTG 1478
      Db 368 CTGCAAGAAAGTCGTTGAGAGCGCCCTCACCAGAAAGATGAGGAGGCCATCGAGGAGCTG 309
      QY 1479 AGCACCGGCATCGTTAAACAAGCTCTCCATGCGCCGCTGCAGACCTGAGGTGCACGCG 1538
      Db 308 AGCACCGGCATCGTTAAACAAGCTCTCCATGCGCCGCTGCAGACCTGAGGTGCACGCG 249
      QY 1539 AGCGACAGCGCACCCCTTTCAGAGAGCGCTCGAGAACATGCACGCCCTCAACCGGATGTC 1598
      Db 248 AGCGACAGCGCACCCCTTTCAGAGAGCGCTCGAGAACATGCACGCCCTCAACCGGATGTC 189
      QY 1599 AGCCTCGACATGGAGAAGCGCATCATCGAGCAGAAAGATCAAGGCCAAGGTGAGAGACA 1658
      Db 188 AGCCTCGACATGGAGAAGCGCATCATCGAGCAGAAAGATCAAGGCCAAGGTGAGAGACA 129
      QY 1659 CAAAACTGAGGCGCAGGAAGCAATTTTTCACACCATTTATCTATATATATAGCGTCTCCA 1718
      Db 128 CAAAACTGAGGCGCAGGAAGCAATTTTTCACACCATTTATCTATATATATATAGCGTCTCCA 69
      QY 1719 ATCTCATTCCTATTTTTCCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1778
      Db 68 ATCTCATTCCTATTTTTCCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9
      QY 1779 TTAAGTGT 1786
      Db 8 TTAAGTGT 1
  RESULT 15
  CB685474
  LOCUS
  DEFINITION
    OSUNEF15004.f OSUNEF Oryza sativa (japonica cultivar-group) cDNA
    clone OSUNEF15004 5', mRNA sequence.
  ACCESSION
    CB685474

VERSION
  CB685474.1 GI:29689199
EST.
  Oryza sativa (japonica cultivar-group)
  Oryza sativa (japonica cultivar-group)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
  1 (bases 1 to 733)
  Jantauriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
  Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
  Large-scale identification of ESTs involved in the interaction
  between rice and Magnaporthe grisea
  Unpublished
  Contact: Rod Ming
  Arizona Genomics Institute
  University of Arizona
  Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
  85721-0088, USA
  Tel: 520 626 3967
  Fax: 520 621 9288
  Email: http://genome.arizona.edu
  PCR Primers
  FORWARD: gta aaa cga cgg cca gtc
  BACKWARD: gga aac agc tat gac cat g
  Plate: 15 row: 0 column: 04
  Seq primer: gta aaa cga cgg cca gtc.
  location/Qualifiers
    1..733
      /organism="Oryza sativa (japonica cultivar-group)"
      /mol_type="mRNA"
      /cultivar="Nipponbare"
      /db_xref="taxon:39947"
      /clone="OSUNEF15004"
      /tissue_type="Leaf"
      /dev_stage="3 week"
      /lab_host="DH10B"
      /clone_lib="OSUNEF"
      /note="Vector: pbluescript II KS +; Site_1: EcoRI; Site_2:
      XhoI; Uninfected Control."
    BASE COUNT      193 a 140 c 209 g 191 t
    ORIGIN
      Query Match      30.4%; Score 584.2; DB 14; Length 733;
      Best Local Similarity 87.3%; Pred. No. 1e-67;
      Matches 640; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
      QY 409 AACTCAGCTAGCCTGAACCATATTGAAGAGCGCTGCTTCTTAGTACCTGTAAAGATGG 468
      Db 1 AACTCAGCACCCTTCAATCATATTGAAGAGCGCTGCTTCTTAGTACCTGTAAAGATGG 60
      QY 469 AATTTATGTGTCGCCCTATCATGGAACCGTGGTATCAGAGAAGTAGTGAGTGGATGT 528
      Db 61 AATTTATGTGTCGCCCTTATCGTGAACCGTGGGATTTAGAGAAGTGTGAGTGGATGT 120
      QY 529 CGAAGAAAAGTGGTATTTCCCGCTTCCGAGCTTAGGGAGACCTGTTCATCTTGGCAACA 588
      Db 121 CAAAGAAAAGTGGAAATCCCTGCTTCTGAGCTCAGGAGCATCTATTATGTTGGGTGACA 180
      QY 589 GTGATGCCACACGCCCATCTGTTTGTGAGGTGTGAGTGGCCCTTGAATCTTTTGTGTTCTCGGTG 648
      Db 181 GTGATGCCACACGCCCATCTGTTTGTGAGGTGTGAGTGGCCCTTGAATCTTTTGTGTTCTCGGTG 240
      QY 649 AAGGACAAATCTTGTCTCAGGTTAAACAAGTTGTGAGGAGTGGAGACAGACAGTGGAGGCT 708
      Db 241 AAGGCAAAATCTTGTCTCAAGTTTAAACAAGTTGTGAGGAGTGGGCAAAAAGTGGAGGCT 300
      QY 709 TGGGAAAAGAACATCGATAGGATGTTCAAGGATGCAATCAGCTGCGGAAAGCGTCTCCGCA 768
      Db 301 TGGGAAAAGAACATCGATAGGATGTTCAAGGATGCAATCAGCTGCGGAAAGCGTCTCCGCT 360
      QY 769 GCGAGACCAACATATCATCTGTTGCTGCTGTTTCTGTGAGTTCAGCGCGCGTGTGAATGCGCCC 828
      Db 361 GCGAGACTAACATATCATCAGGTGCTGCTCTCTGTGAGTTCAGCTGCGAGTTGAATGCGCCT 420

```

Qy	829	TGATGAAGCTTCCGAAGTCTGAAGCACTGTCAAGCTAGGATGCTTCTGATTGGTGGCTGTA	888
Db	421	TGATGAAGCTTCCAAAGTCGGAATGCCTATCTGCTAGGATGCTGTTGATTGGTGGCA	480
Qy	889	AAATGGGAAGCTAGTATCAAAATCTGGTTGCCAAAGGATGCAAGAGGTTGTTGG	948
Db	481	AGATGGGAAGTTGGTGGTTAAACATTTGATTGCCAAGGATGCAAGAAAGTTGTTGG	540
Qy	949	TGAACCGCTCCGTGGAAAGGTTGGATGCTATTCGTGAGGATGAAAGATATAGAGTCG	1008
Db	541	TGAACCGTTCAGTGGAAAGGTTGGATGCCATCCGGAAGAGATGAAGACATTGAGATTG	600
Qy	1009	TGTACAGGCTCTCTCAGACATGTATCAAGCTGTGCTGAAGCTGATGCTGTTTCA	1068
Db	601	TGTACAGGCTCTTACAGAGATGTATGAAGCGCTGCCGAAGCTGATGCTGTTTCA	660
Qy	1069	GCACGGATCTGAACCTTCTTTCGCAAGAGAACACGACAGGCACTCCCCCTGTCT	1128
Db	661	GCACGGATCCGAACCCCATTTGTTTCAAAAGGAGACGACAGGCGCTTCCCGCTATT	720
Qy	1129	CTGATACTATGGG	1141
Db	721	CTGATGCTATGGG	733

Search completed: December 15, 2003, 16:07:25  
Job time : 4199 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	1324	100.0	1924	22	AAF54814
2	1196.4	62.2	1778	22	AAF54816
3	1107	57.5	1480	22	AAF54822
4	710.4	36.9	2055	22	AAF54818
5	679.6	35.3	1632	24	ABZ13396
6	589.2	30.6	1983	22	AAF54820
7	223.2	11.6	510	22	AAF54815
8	179.8	9.3	312	22	AAF54813

XX WPI; 2001-159865/16.  
DR P-PSDB; AAB31960.  
DR

PT New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme  
PT useful for producing large amounts of the encoded polypeptides used in  
PT screening compounds for potential herbicidal activity -

PS Claim 2; Page 47-48; 77pp; English.

The present sequence encodes a Glu-tRNA reductase, which is an aminolevulinic acid biosynthetic enzyme. The enzyme converts Glu-tRNA-Glu to glutamate 1-semialdehyde (GSA) with the concomitant release of tRNA-Glu. GSA aminotransferase then converts GSA to 5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs and genes encoding homologous proteins from the same or other plant species, for creating transgenic plants in which the polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found, for overexpression in bacterial or yeast hosts to efficiently produce large amounts of the encoded polypeptides which could then be used for screening different compounds for potential herbicidal activity, and as hybridisation probes and amplification primers.

SQ Sequence 1924 BP; 504 A; 465 C; 551 G; 404 T; 0 other;

```
Query Match      100.0%; Score 1924; DB 22; Length 1924;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1924; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	CGAGCGTCCGATCAATAAAGAGAGCTTTGGAAAGTTGCCAAGGCTTCGATTTTCGCTA	60
Db	1	CCAGCGTCCGATCAATAAAGAGAGCTTTGGAAAGTTGCCAAGGCTTCGATTTTCGCTA	60
Qy	61	ATCGACGATATATGGCGACACAGCGTCAGCGACACACCGCGCGCAGCAGCCGCCACCA	120
Db	61	ATCGACGATATATGGCGACACAGCGTCAGCGACACACCGCGCGCGCAGCAGCCGCCACCA	120
Qy	121	CGCCAAAGCGCGGGGGTTCGTTCGGGCCCTTCGCAGAGGGTCCGCGCGCGGCGAGGC	180
Db	121	CGCCAAAGCGCGGGGGTTCGTTCGGGCCCTTCGCAGAGGGTCCGCGCGCGGCGAGGC	180
Qy	181	GGCGTCCGGGGTGTGGGTGCGACGCGCGCGCGGTGGAGGCCAGCGCAGCGCGTGG	240
Db	181	GGCGTCCGGGGTGTGGGTGCGACGCGCGCGCGGTGGAGGCCAGCGCAGCGCGTGG	240
Qy	241	CAAAGGCTCCGACGCTCGCGGCCCTCGAGCAGTTCAGAGATACCGCGACCGGTACATGA	300
Db	241	CAAAGGCTCCGACGCTCGCGGCCCTCGAGCAGTTCAGAGATACCGCGACCGGTACATGA	300
Qy	301	AGAAAGAGATACCATAGCTGTGATAGGCCTCAGTGTCACACAGCACCAAGTGAGATGC	360
Db	301	AGAAAGAGATACCATAGCTGTGATAGGCCTCAGTGTCACACAGCACCAAGTGAGATGC	360
Qy	361	GTGAAAACTCTCGTTGCAGAGGAACGTGCGCCCGTCTATTTCAGAACTCACTAGCC	420
Db	361	GTGAAAACTCTCGTTGCAGAGGAACGTGCGCCCGTCTATTTCAGAACTCACTAGCC	420
Qy	421	TGAACCATATGAAGAGGCTCGTGTCTTTAGTACTCTGTAATAGAAATGGAATTTATGTG	480
Db	421	TGAACCATATGAAGAGGCTCGTGTCTTTAGTACTCTGTAATAGAAATGGAATTTATGTG	480
Qy	481	TGGCGCTATCATGGAAACGTGTGATCAGAGAAGTAGTGGAATGGAATGGAAGAAAGTG	540
Db	481	TGGCGCTATCATGGAAACGTGTGATCAGAGAAGTAGTGGAATGGAATGGAAGAAAGTG	540
Qy	541	GTATTCGCCGCTTCAGAGCTTAGGAGCACCTGTTTCATCTTCGGAAGCAGGTATGCCAC	600
Db	541	GTATTCGCCGCTTCAGAGCTTAGGAGCACCTGTTTCATCTTCGGAAGCAGGTATGCCAC	600
Qy	601	GCCATCTGTTGAGGTGTCACTGGCCCTTGACTCTTTGGTTCCTCGGTGAAGGACAAATCC	660







```

Db 606 TGGTTGTCAACATTTGATTGCCAAGGATGCAAGAGGTTCTGTGGTGAACCGTTCTG 665
QY 961 TCGAAGGCGTGCATCTATTCTGTGAGGATGAAGATATAGATCGTGTACAGCCCTC 1020
Db 666 TGGAAAGGTTGATGTCATTCGCCAAGAGATGAAGATATTGAGATTGTGTACAGCCCTC 725
QY 1021 TCTCAGACATGTATCAAGCTGTCTGTGAAGCTGTATGCTGTTCACAGCAGCGATCTG 1080
Db 726 TTACAGAGATGTATGAAGCGCTGTGAAGCTGTATGCTGTTCACAGCAGCGATCTG 785
QY 1081 AAATTCATTTGCGAAAGAACACGACAGAGGCTATCCCTCTCTCTGTATCTATG 1140
Db 786 AATCTCTATTATTCAGAAAGAGCATGACAGAGCGCTTCTCTCTATTTCTGTCTGTG 845
QY 1141 GAGGTGTTCCGCTGTTTGTGACATATCTGTCCAGGAATCTCAGCGATGTGTCTG 1200
Db 846 GTGGTGTTCGGCTTTTGTGACATATCTGTCCAGGAATGTGCGTGCCTGTGTATCTG 905
QY 1201 AAGTTGGCGCTGCACGAGTGTACAATGTTCGACGACTTGAAGAGGTGGTGAAGCAACA 1260
Db 906 AGTTCGAGCATGCACGGGTATACATGTTCGACGACTTGAAGAGGTGGTGAAGCAACA 965
QY 1261 AGAGAGACCGGCTCAGGAAAGCAATGGAGGCGACGACATCATCCGAAAGTGAAGAC 1320
Db 966 AGGAAGACCGTGTGAGGAAAGCAATGGAGGCGCCAAACAATCATATTACCAGAACTGAAC 1025
QY 1321 GGTTCGAGGATGGAGGACTCGCTGGAGACCGTTCGACCATCAAGAGCTGAGTGTGT 1380
Db 1026 GGTTCGAGGATGGAGGACTCACTGGAGACCGTTCGACCATCAAGAGCTGAGTGTGT 1085
QY 1381 ACGCGACAGGATCAGGCGCTCGAGCTCGAAGTGCCTGCAGAAAGTAGGTGAGGACG 1440
Db 1086 ACGCGACAGGATCAGGCGCATCCGAGCTCGAAGTGTCTGAGAAGATCGGGGAAGACA 1145
QY 1441 CCTCACCAGAAAGATGAGGAGCGCATCGAGGAGCTGAGCACCGGCTATCGTTAAAGC 1500
Db 1146 ATCTCAACAAGAAAGATGAAGGTCCATCGAGGAGCTGAGCACCGGCTATGTAACAAGC 1205
QY 1501 TCCTCATGCGCGCTGACGACCTGAGTGGCGAGCGGACGACGACGCGCCCTTGAGC 1560
Db 1206 TCCTTCACGGCCCATCTGACGACCTGAGATGCGAGCGGACGACGCGCCCTTGAGC 1265
QY 1561 AGACGCTCGAAGAACATGCACGCGCTCAACCGGATGTTGAGCCTCGACATGGAAGAGCGA 1620
Db 1266 AAACGCTTGAGAAACATGCACGCGCTCAACAGAAATGTTCAACCTCGACGAGAGCGG 1325
QY 1621 TCATCGACGAGAGATCAAGGCCAAGGTGGAGAGACACAAACTGAGGCCAGGAAGCAA 1680
Db 1326 TCCTTGAGCAGAAAGATCAAGGCCAAGGTAGAGAGACCCAAAGCTGAGACGAGAGAC 1385
QY 1681 TTTTCTTACCACTTATCTATATATATAGCGTCTCCCAATCTCATTCATTTT 1733
Db 1386 TTGCCCGTC--TGATATCTACTTATATCTGCTCCAGGAATGTGCTATCTCT 1436

```

## RESULT 4

AAFP54818  
ID AAFP54818 standard; cDNA; 2055 BP.

AC AAFP54818;

DT 15-MAY-2001 (first entry)

DE Nucleotide sequence of a soybean Glu-tRNA reductase enzyme.

KW Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;  
tetrapyrrolic pigment; chlorophyll; transgenic plant; ss.

OS Glycine max.

FH Location/Qualifiers  
FT CDS  
FT CDS  
/\*tag= a

/product= "Glu-tRNA reductase"

WO200109304-A2.

08-FEB-2001.

28-JUL-2000; 2000WO-US21008.

30-JUL-1999; 99US-0146600.

(DUPO ) DU PONT DE NEMOURS & CO E I.

Cahoon RE, Gutteridge S, Harvell LT, Rafalski JA, Tao Y, Weng Z;

WPI; 2001-159865/16.

P-PSDB; AAB31964.

New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme  
useful for producing large amounts of the encoded polypeptides used in  
screening compounds for potential herbicidal activity -

Claim 2; Page 55-57; 77pp; English.

The present sequence encodes a Glu-tRNA reductase, which is an  
aminolevulinic acid biosynthetic enzyme. The enzyme converts  
Glu-tRNA-Glu to glutamate 1-semialdehyde (GSA) with the concomitant  
release of tRNA-Glu. GSA aminotransferase then converts GSA to  
5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of  
tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding  
aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs  
and genes encoding homologous proteins from the same or other plant  
species, for creating transgenic plants in which the polypeptides are  
present at higher or lower levels than normal or in cell types or  
developmental stages in which they are not normally found, for  
overexpression in bacterial or yeast hosts to efficiently produce  
large amounts of the encoded polypeptides which could then be used for  
screening different compounds for potential herbicidal activity, and  
as hybridisation probes and amplification primers.

Sequence 2055 BP; 596 A; 419 C; 465 G; 575 T; 0 other;

Query Match

Best Local Similarity 36.9%; Score 710.4; DB 22; Length 2055;

Matches 1019; Conservative 0; Mismatches 471; Indels 6; Gaps 2;

QY 189 GGGGTGTGCGGTGCGACGCGCGCGGTGGAGGCCAGCGCGCGTGGCAAGGCT 248

Db 252 GGGGTATTCTGCTGCGACGCTCAGCCCTCTGATGATCATCTGTTGCTCCAAATATGCC 311

QY 249 GCCAGCTGCGCGCGCTCGACGAGTTCAAGATAT---CGCGGACCGGTACATGAAGGAA 305

Db 312 ACCGCTCTCTCCGCTCTTTGAGCAGCTCAAGACTTCTGCAGCTGATAGATATACAAAGGAA 371

QY 306 AGGAGTACCATAGCTGTGATAGGCTCTAGTGTACACAGCAGCAGCTGAGATCGGTGAA 365

Db 372 AGAGCAGCATATTCGCCATTTGGGCTAGTGTGACACTGACCTGTGGAATGCGTGA 431

QY 366 AAACCTGTGTTGCAGAGGAACTGTGGCCCGTGTCTATTCAAGAACTCACTAGCTGAA 425

Db 432 AAACCTGTGTTGCAGAGGAACTGTGGCCCGTGTCTATTCAAGAACTCACTAGCTGAA 491

QY 426 CATATTGAGAGGCTGTGTTCTTAGTACCTGTATAGAAATGGAATTTATGTGTGGG 485

Db 492 CATATTGAGAGGAGCAGCTGTTCTGAGTACCTGCAATCGAATGGAGATATATGTTCTGCC 551

QY 486 CTATCATGGAACCGTGGTATCAGAGAACTAGTGTGACTGTGCAAGAAAGTGTGTTAT 545

Db 552 CTGTCCCAACATCTGTGTCTCAAGAACTAGTGTGCAATGGAGATATATGTTCTGTC 611

QY 546 CCCGCTTCGAGCTTAGGAGCAGCTGTTCATCTTTGCAAGCAGTGTGACACAGCCAT 605

Db 612 CCTGTTTCAGAGCTTAGCCAGCAGCGGTTTTTACTTTTACACATGATGCCACAGCAT 671

```
QY 606 CTGTTTGAAGTGTACGCTGCTTGAATCTTTGGTTCTCGTGAAGGACAAATCCTGCT 665
Db 672 CTTTTTGAAGTATCAGCAGGCTTGAATCTTTGGTTCTCGTGAAGGACAAATCCTTCT 731
QY 666 CAGGTTAAACAAAGTTGTAGAGGAGTGACAGAACAGTGGAGGCTTGGGAAGAACATCGAT 725
Db 732 CAGGTTAAGCAAGTTGTTAAAGTTGACCAAGAGTTAACGGCTTTGGGAGAAATATCAGT 791
QY 726 AGGATGTTCAAGATGCAATCATCTGTGGAAGCGTGTCCGAGCGAGACCAACATATCA 785
Db 792 GGGCTATTCAAGCATGCAATTAATCTGTGCGGAAAGGGTTAGAACTGAGACTAATATGCT 851
QY 786 TCTGGTGTCTTCTGTCTGATTCAGCGCGGTTTGAATCGCGCTGATGAAGCTTCCGAG 845
Db 852 TCTGGGCGAGTTTCTGTGAGCTCAGCTGCGTGTGAGTTGGCTTATATGAAGTTACTGAA 911
QY 846 TCTGAAGCACTGTCACTAGTAGTCTTCTGATTGTTGCTGTGTAATAATGGAAAGCTAGTG 905
Db 912 GCCTCACAGATAATGCCAGGATGTTGGTTATTGGTGTGCGCAAGATGGAAAGCTTGTG 971
QY 906 ATCAACATCTGGTTGCCAAAGATGCAAGAGGTTGTTGTGTGAACCGCTCCGTGGA 965
Db 972 ATCAACATCTGGTGGCAAGGTTGCAAAAAGATGGTGGTTGTCAATAGAACTGAGGAG 1031
QY 966 AGGGTGGATGCTATTCTGTGAGGAGATGAAGATATAGAGATCGTGTACAGGCTCTCTCA 1025
Db 1032 AGAGTTGCTGCAATAGCTGAAGACTGAAGGATATTGAGATTATCTACAAACCCCTTCA 1091
QY 1026 GACATGATCAAGCTGCTCTGAAGCTGATGTGTTTCCAGCAGCACCGCATCTGAAACT 1085
Db 1092 GAAATGCTCACCTGTGCGCAAGCAGATTTAGTTTCCACCAGTACTGCATCAGAAAC 1151
QY 1086 TCATTGTTGCAAAAGAACACGAGGACCTCCCGCTGTCTGTGATCTATGGAGGT 1145
Db 1152 CCATTATTCTGAAAGAAATGTCGAAGGACCTCTCTCTGCAAGTCAAGAACTGGAGC 1211
QY 1146 GTTCGCTGTTGTGACATATCTGTCCCGAGGATGTGAGCGCATGTGTCTGAAGTT 1205
Db 1212 CGTCGCTTTTTCATTGATATCTGTGTTCCCGGATGTGGTTCAATGTCTCAGACCTT 1271
QY 1206 GCGCTGCAAGATGTACAAATGTGACGACTTGAAGAGGTTGGTGAAGCCCAACAGGAG 1265
Db 1272 GAGTCTGTGCGAGTTTACAAATGTGACGACCTTAAAGAGGTTGTGCTGCCAATAAAG 1331
QY 1266 GACCGCTCAGAAAGCAATGAGGCGCAGACAAATCATCACGAGAACTGAGACGGTTC 1325
Db 1332 GATCGCTAAGAAAGCAATGGAAGCAGGCAATCATTTGCTGAAGAACTTAAGCAATTC 1391
QY 1326 GAGCATGAGGAGCTCGCTGGAGACCGTTCCGACCATCAAGAAGCTGAGGTCTGACGG 1385
Db 1392 GAAGCTTGAGGAGCTCACTGGAAACTGTTCTTACTATTAAAGAAATGAGGCTTATGCT 1451
QY 1386 GACAGATCAGGCTCGAGCTCGAAGATGCTGCGAAGATGAGTGAAGACCCCTTC 1445
Db 1452 GAAAGATCAGGCTTGTCTGAGTGTGAGAAAGTCTTGAAGATGGGTGATCA---TATA 1508
QY 1446 ACCAAGATCAGGAGAGCCATCGAGGAGCTGAGCACCGCATCGTTAAACAGCTCTTC 1505
Db 1509 CCAAGAAACCGGAGAGCTGTGATGACCTTATGTCGGGGTATAGTGAATAGTTGCTT 1568
QY 1506 CATGCGCCCTGCAGACCTGAGGTGCGACGCGCAGCAGCGCACCGCACCCCTTGACGAGC 1565
Db 1569 CATGTGCCAATGCAATTTAAGTTGTGATGGAGACGACGCGGACTCTTAGTAGTGACA 1628
QY 1566 CTCGAGAACATGACCGCCCTCAACCGGATGTTTCAGCTCGCATGAGGAGCGGATCATC 1625
Db 1629 CTGGAGAACATGAATGCTTTGAATAGGATGTTCAACCTTTGAGACAGAAATATCTGTTG 1688
QY 1626 GAGCAGAAATCAAGCCCAAGTGGAGAGACACAAAACCTGAGGCCAGGACCAAT 1681
Db 1689 GAGGAGAGATTCGAGCAAGGTCGAAACAAACCAAGAAATGAATTAACACCAAT 1744
```

```
RESULT 5
ABZ13396
ID ABZ13396 standard; DNA; 1632 BP.
XX
AC ABZ13396;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1201.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX
PA (SCEI ) SCRIPPS RES INST.
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT .
XX
PS Claim 144; SEQ ID NO 1201; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 1632 BP; 463 A; 289 C; 428 G; 452 T; 0 other;
```

```
Query Match 35.3%; Score 679.6; DB 24; Length 1632;
Best Local Similarity 68.6%; Pred. No. 2.3e-137;
Matches 985; Conservative 0; Mismatches 439; Indels 12; Gaps 3;
QY 235 CCGTGGCAAGAGCTGCCAGCGTCCGCGCTCGAGCAGTTCAGATAT---CCGCGGACC 291
Db 206 CTGGCTCTAATGCGAGTAGCATCTGCTCTTGACAACTCAAGAACTCTGCGAGTGCATC 265
QY 292 GGTACATGAAGAAAGGAGTACCATAGCTGTGATAGGCTCAGTGTATACACAGACACAG 351
Db 266 GATATACAAAGAAAGAGAGAGTATTTGTGTGATGGACTTAGTATTACACAGCTCTG 325
QY 352 TGGAGATCGGTGAAGAACTTCTGTCAGAGGAACTGTGGCCCGCTGCTATTCAAGAAC 411
Db 326 TTGAGATCGGTGAAGAGCTTCTATACCTGAAGCTGAATGGCCAAAGAGCTATTGCTGAAT 385
QY 412 TCACCTAGCTGAACCATATTGAAGAGGCTGCTGTTCTTTAGTACCTGTATAGAAATGGA 471
Db 386 TGTGTGTTGAATCATATTGAAGAGCTGCTGTCTGTAGTACTTGTAAACGATGGAGA 445
```

QY 472 TTTATGTCGTGGCGCTATCATGGAACCGTGGTATCAGAGAGTACTGGACTCGATGTGCA 531  
 Db |||||  
 QY 446 TTTATGTCGTGGCGCTATCATGGAACCGTGGTATCAGAGAGTACTGGACTCGATGTGCA 505  
 Db |||||  
 QY 532 AGAAAGTGGTATTCCTCGCTTCGAGCTTAGGGAGCACCTGTTTCATCTTTCGCAAGCAGTG 591  
 Db |||||  
 QY 506 AGACGAGTGGATCCAGTTTCAGAGATTTGTCAACACCGTTTCTGTTGTACACAGG 565  
 Db |||||  
 QY 592 ATGCCACACCGCATCTGTTTGGAGTGTAGCTGGCTTGAATCTTTCGTTTTCGTTGTAAG 651  
 Db |||||  
 QY 566 ATGCTACACAGCATATATTTGAAGTCTCAGCTGGTCTGACTCTCTGTTCTCAGAGAG 625  
 Db |||||  
 QY 652 GACAAATCTGCTCAGGTTAAACAAAGTTGTGAGAGTGGACAGACAGTGGAGCTTGG 711  
 Db |||||  
 QY 626 GTCAGATCTTGCACAGGTGAACAAAGTTGTGAAGTTGGTCAAGAGTGAATGGCTTGG 685  
 Db |||||  
 QY 712 GAAGAATCATGATAGGATGTTCAAGGATGCAATCACTGCTGGAAAGCGTTCGCGAGG 771  
 Db |||||  
 QY 686 GGAGGAATATCAGCGGCTGTTTAAACACGCGATAACTGTTGTAAGCGTGTAGAACAG 745  
 Db |||||  
 QY 772 AGACCAACATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831  
 Db |||||  
 QY 746 AGACAAATATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 805  
 Db |||||  
 QY 832 TGAAGCTTCGAGAGTCTGAAGCACTGCTCAGCTAGGATGCTCTGATGCTGCTGCTGCTG 891  
 Db |||||  
 QY 806 TGAAGCTTCGAGAGTCTGAAGCACTGCTCAGCTAGGATGCTCTGATGCTGCTGCTGCTG 865  
 Db |||||  
 QY 892 TGGGAAGCTAGTATCAACATCTGTTTGCACAAAGGATGCAAGAGGTTGTTGCTGGA 951  
 Db |||||  
 QY 866 TGGGAAGCTAGTATCAACATCTGTTTGCACAAAGGATGCAAGAGGTTGTTGCTGGA 925  
 Db |||||  
 QY 952 ACCGCTCGTGAAGAGGTTGATGCTATCTGCTGAGGATGCAAGATATAGAGATCGGT 1011  
 Db |||||  
 QY 926 ACAGAGTGAAGAGAGGTTTTCAGCTATCCGTTGAGGATGCTGCTGCTGCTGCTGCTG 985  
 Db |||||  
 QY 1012 ACAGGCTCTCTCAGACATGATCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1071  
 Db |||||  
 QY 986 ATGACCTCTGATGAGATGCTAGCTTGTGCTTCAAGAGCGGATGTTGTTTACAGCA 1045  
 Db |||||  
 QY 1072 CCGCATCTGAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1131  
 Db |||||  
 QY 1046 CAGCTCTGAGAGCGCATTTCTTAAAGGAGCATGTAGAGATCTCCTCAAGCTTCTC 1105  
 Db |||||  
 QY 1132 ATACTATGGAGGTTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1191  
 Db |||||  
 QY 1106 CAGAGTTTGAAGGAGTGAAGGCTTTTGTGATCTCTGTTCCGAGAAATGTTGATCTT 1165  
 Db |||||  
 QY 1192 GTGTGCTCTGAAGTTGGCGCTGACGAGTGTAAATGTCGAGCACTTGAAGAGGTTGG 1251  
 Db |||||  
 QY 1166 GTGTGCTCTGAAGTTGGCGCTGACGAGTGTAAATGTCGAGCACTTGAAGAGGTTGG 1225  
 Db |||||  
 QY 1252 AAGCCAAAGAGAGGAGCGGCTCAGAAAGCAATGGAGGCGCAGACATCATCACCGAAG 1311  
 Db |||||  
 QY 1226 CTGCCAAAGAGAGGAGGATGAGAAAGCAATGGAAGCTCAAAACATATCATCAGAGG 1285  
 Db |||||  
 QY 1312 AACTGAGACGTTGAGGAGTGAAGGAGTCTGCTGAGACCGTTCCGACCATCAAGAGC 1371  
 Db |||||  
 QY 1286 AATCCACAGTTTGAAGCATGAGGAGGATTCATTGGAGACCGTTCCACGATCAAGAGT 1345  
 Db |||||  
 QY 1372 TGAGTCTGACCGGACAGGATCAGGGCTCGAGCTCGAGAGTGTGCTGCTGAGAGGATG 1431  
 Db |||||  
 QY 1346 TAAGAGCTTATGACAGAGATCAGATGGGAGAGCTCGAGAGTGTGCTGCTGAGAGGATG 1405  
 Db |||||  
 QY 1432 GTGAGGAGCGCTTCAACAGAGATGAGAGAGCCATCGAGGAGTGTGCTGCTGAGAGGATG 1491  
 Db |||||  
 QY 1406 GAGATGA---TATCAACAG 1462  
 Db |||||  
 QY 1492 TTAACAGCTCTCTCATGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551  
 Db |||||  
 QY 1463 TGAACAGATCTTACATGCTTCAATGAGGATTTGAGATGTTGAGAGAGAGAGAGAGAGAGAG 1522  
 Db |||||

QY 1552 CCCTTGACGAGCGCTCGAGACATGCGCCCTCAACCGGATGTTTCAGCTCGCATGG 1611  
 Db |||||  
 QY 1523 CGCTGAGGAGACCTTGCATATGCTGCTTGAACAGAAATGTACCGTC-----TTG 1576  
 Db |||||  
 QY 1612 AGAAGCGCATCATCAGCAGAGAGATCAAGGCGCAAGTGGAGAGACACAAAACCTGA 1667  
 Db |||||  
 QY 1577 AGAAGAGACATTTTGGAGGAAAGCTTAAGGCAATGCGGAAACACACAGAACTAA 1632  
 Db |||||

RESULT 6

AAF54820  
 ID AAF54820 standard; cDNA; 1983 BP.

XX AAF54820;

XX AC AAF54820;

XX DT 15-MAY-2001 (first entry)

XX Nucleotide sequence of a soybean Glu-tRNA reductase enzyme.

XX Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;

XX tetrapyrrolic pigment; chlorophyll; transgenic plant; ss.

XX Glycine max.

XX Key Location/Qualifiers

XX CDS 289..1693

XX FT /\*tag= a

XX FT /product= "Glu-tRNA reductase"

XX WO200109304-A2.

XX PN 08-FEB-2001.

XX PD 28-JUL-2000; 2000WO-US211008.

XX PF 30-JUL-1999; 99US-0146600.

XX PR (DUPO ) DU PONT DE NEMOURS & CO E I.

XX PA Cahoon RE, Gutteridge S, Harvell LT, Rafalski JA, Tao Y, Meng Z;

XX PI WPI, 2001-159865/16.

XX DR P-PSDB; AAB31966.

XX PT New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme

XX PT useful for producing large amounts of the encoded polypeptides used in

XX PT screening compounds for potential herbicidal activity

XX PS Claim 2; Page 62; 77pp; English.

XX The present sequence encodes a Glu-tRNA reductase, which is an  
 CC aminolevulinic acid biosynthetic enzyme. The enzyme converts  
 CC Glu-tRNA-Glu to glutamate 1-semialdehyde (GSA) with the concomitant  
 CC release of tRNA-Glu. GSA aminotransferase then converts GSA to  
 CC 5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of  
 CC tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding  
 CC aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs  
 CC and genes encoding homologous proteins from the same or other plant  
 CC species, for creating transgenic plants in which the polypeptides are  
 CC present at higher or lower levels than normal or in cell types or  
 CC developmental stages in which they are not normally found, for  
 CC overexpression in bacterial or yeast hosts to efficiently produce  
 CC large amounts of the encoded polypeptides which could then be used for  
 CC screening different compounds for potential herbicidal activity, and  
 CC as hybridisation probes and amplification primers.

XX Query 1983 BP; 514 A; 413 C; 479 G; 577 T; 0 other;

Query Match

Best Local Similarity 30.6%; Score 589.2; DB 22; Length 1983;

Matches 902; Conservative 0; Mismatches 473; Indels 9; Gaps 2;

QY 282 TCGCCGACCGGTACATGAAGAGAGAGTACCATTGCTGTGATAGCCCTCAGTGTACAC 341

```

|||||
317 TCCGCCGACAGATATACGAGGAAAGAGTTGCAATATTTGCTATAGGCTGACATTCAC 376
QY
342 ACAGACCCAGTGGAGATCGGTGAAAAAATTCGCTGTTGAGAGGAACTGTGCCCCGTGCT 401
Db
377 ACTGCTCCCGTTGAGATCGTGAAGAGCTTGCAATTCAGAAATCCCATTTGGGCTCAGGCT 436
QY
402 ATTCAGAACTCCTAGCTGCAACCATATTTGAAGAGGCTGCTGTTCTTAGTACCTGTAAT 461
Db
437 ATTAAGGACCTTTGGCTTTGAAACCATATCGAAGAGCCGGTTCTCAGCAGCTGTAAAC 496
QY
462 AGAATGGAATTTATGCTGGCGCTATCATGGAAACCGTGTATCAGAGAAAGTAGTGAC 521
Db
497 CGCATGGAGATCTATGTTGGCTCTTTCCAGCACCCGTTGTTTAAAGAAAGTTACTGAT 556
QY
522 TGGATGTCGAAGAAAGTGTATTCGCCGTTCCGAGCTTAGGAGCACCTGTTCACTCTTG 581
Db
557 TGGATGTCGAAGAGTGAAGCGGATTTCAATACCTGAGCTTTGTGAGCACCAAGTTTGGCTG 616
QY
582 CGAAGCAGTGTATGCCACACGCCATCTGTTGAGGTGTCAGCTGGCGCTTGTACTCTTTGGTT 641
Db
617 TATAAGCGGATGTCACGACGATCTCTTTGAGTGGCGGAGGCTTGACTCACTTGT 676
QY
642 CTCGCTGAAGGACAAATCCTTGCTCAGGTTAAACAAAGTTGTGAGGAGTGGACAGAACAGT 701
Db
677 CTGGGGAAGGTCAAATTTCTGCTCAGGTGAAGCAGGTTGTGAAAGCTGGACAGGAGTG 736
QY
702 GGAGCTTTGGGAAGAAACATCGATAGGATGTTCAAGGATGCAATCACTGCTGGAAAGCGT 761
Db
737 CCTGGTTTGTATGAAGAAATTAGTGGTTTGTTCAGCAGGCGATCTCGGTTGGGAAGCGG 796
QY
762 GTCGCGACGACGACCAACATATCATCTGCTGCTGTTCTGCTGAGTTCAGTCGCGGCTTGAA 821
Db
797 GTTAGAATCTGAGACTAAATTTCTGCTGAGTGGTTCTGCTGAGCTGCGCTGCTGTGAG 856
QY
822 CTGGCCTGTAGTGAAGCTTCCGAAAGTCTGAAGCACTGTGAGTAGATGCTTCTGATTGGT 881
Db
857 CTCGCACTGATGAAGCTTCCGGAATTCCTCTTCTGCTGATTTCTGGAGTGTGCTGTTGGT 916
QY
882 GCTGTAAATGGGAAGCTAGTATCAAACTCTGTTGTCGAGTTCGAAAGGATGCAAGAGTT 941
Db
917 GCAGGGAAGATGGGAAGCTTGTATTAAGCATTTTGGCTGCGCAAGGGTGAGAAAGATG 976
QY
942 GTTGTGTGAACCGCTCCGCTGGAAAGGTTGATGCTATTCGTGAGGAGATCAAGATATA 1001
Db
977 GTTGTGTTTAAACAGACTGAAGAGAAAGTTATGCCATTCGAAAGAGTTGAAGATGTT 1036
QY
1002 GAGATCGTGTACAGCTCTCTCAGACATGTATCAAGCTGCTGTGAGCTGATGTCTGTG 1061
Db
1037 GAGATGTTATTTAGACCATTTTCAGATATGCTGGCGTGTGCTGTAAGCTGATGTGATC 1096
QY
1062 TTCACGACGACCGCATCTGAAACTTCATTGTTCCGAAAGAACACGAGAGGCACTCCCC 1121
Db
1097 TTCACGACGACGCTGTGAATCACCATTGTTTCTAAACAGAAATGTCAGATGCTTCTCT 1156
QY
1122 CCTGCTCTGTATATGAGGAGTGTTCGCCCTGTTTGTGCAATATCTGCCCCAGGAAT 1181
Db
1157 CTGGT-----TAATCATGGGAGAGGCGGCTTTTGTGATATCTATTCCTAGGAT 1210
QY
1182 GTCAGCGCATGTGTCTGAAGTTGGCGCTGACAGTGTACATGTGACGACTTGAAA 1241
Db
1211 GTGGAACCGGGTGTCTCAGATCTGGAGACTGCACTTGTGTACAATGTGGATGATCTGAAG 1270
QY
1242 GAGGTGTGGAAGCAACAGGAGGACCGGCTCAGGAAGCAATGGAGGCGGACACATC 1301
Db
1271 GAAGTTGTGACGTAAACAGGAGACAGGCTTCAGAAAGCTGAGGAAGCCCGGGAAT 1330
QY
1302 ATCACCGAAGAACTGAGACGTTCCGAGCATGGAGGACTCGCTGGAGACCGTTCCGACC 1361
Db
1331 ATACTAGAGAGTTGAATTAATTCGAAGCTTGGAAAGACTCTCTGGAAACTGTCTCTACT 1390
QY
1362 ATCAGAGCTGAGGTGTACCGCGGACAGGATCAGGCGCTCGGAGCTCGAGAAAGTGCCTG 1421

```

```

Db 1391 ATTAAGAGTTTAGAGCTTATGTTGAGAGGATAAGAGCCTCTGAGATGGAGAGTGT 1450
QY 1422 CAGAAAGTAGTGAGGAGCGCCTCACAAGAAAGATGAGGAGGCAATCGAGAGCTGAGC 1481
Db 1451 TCGAAGATGGGTCTCTGATG---TCTCAAAGCAACAGAAAGATGCAATTTATGCCCTTAGT 1507
QY 1482 ACCGGCATCGTTAAACAGCTCTCCATGCGCCGCTGCAGCACCTGAGGTGGACGGCAGC 1541
Db 1508 ATGGGTATTTGTAATAGCTACTTCTATGTCCTCCATGCAAGCCTTAAGGTGTATGGGAA 1567
QY 1542 GACAGCGCACCCCTTGACGAGCCTCGAGAACATGCAAGCCCTCAACCCGATGTTTCAGC 1601
Db 1568 AATGATAGTAGTCTCTGAGTGAGGTACTTGAAGATATGCGTGCCCTTAACAGAAATGTACGAT 1627
QY 1602 CTCACATGGAAGAGGCGATCATCGAGCAGAAAGATCAAGGCCAAGGTGGAGAAAGACAA 1661
Db 1628 CTTGAGACAGAAATTTCTCTTGTATCGAAGAAAGATCAGAGTCAAGATGGAACGGGTTTCCAG 1687
QY 1662 AACT 1665
Db 1688 AAGT 1691

RESULT 7
AAFS4815
ID AAF54815 standard; cDNA; 510 BP.
XX
AC AAF54815;
XX
DT 15-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a rice Glu-tRNA reductase enzyme.
XX
KW Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;
KW tetrapyrrolic pigment; chlorophyll; transgenic plant; ss.
XX
OS Oryza sativa.
XX
FH Key Location/Qualifiers
FT CDS 9..264
FT FT /*tag= a
FT FT /product= "Glu-tRNA reductase"
FT FT /transl_except= "(pos: 234..237, aa: Xaa)"
FT FT /note= "Xaa is an unknown amino acid"
XX
FN MO200109304-A2.
XX
PD 08-FEB-2001.
XX
PF 28-JUL-2000; 2000WO-US21008.
XX
PR 30-JUL-1999; 99US-0146600.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon RE, Gutteridge S, Harvell LT, Rafalski JA, Tao Y, Weng Z;
XX
DR WPI; 2001-159865/16.
XX
DR P-PSDB; AAB31961.
XX
New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme
XX useful for producing large amounts of the encoded polypeptides used in
XX screening compounds for potential herbicidal activity -
XX
Claim 2; Page 51; 77pp; English.
XX
The present sequence encodes a Glu-tRNA reductase, which is an
XX aminolevulinic acid biosynthetic enzyme. The enzyme converts
XX Glu-tRNA-Glu to glutamate 1-semialdehyde (GSA) with the concomitant
XX release of tRNA-Glu. GSA aminotransferase then converts GSA to
XX 5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of
XX tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding
XX aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs

```

CC and genes encoding homologous proteins from the same or other plant  
 CC species, for creating transgenic plants in which the polypeptides are  
 CC present at higher or lower levels than normal or in cell types or  
 CC developmental stages in which they are not normally found, for  
 CC overexpression in bacterial or yeast hosts to efficiently produce  
 CC large amounts of the encoded polypeptides which could then be used for  
 CC screening different compounds for potential herbicidal activity, and  
 CC as hybridisation probes and amplification primers.

XX  
 SQ Sequence 510 BP; 144 A; 102 C; 130 G; 111 T; 23 other;

Query Match 11.6%; Score 223.2; DB 22; Length 510;  
 Best Local Similarity 70.7%; Pred. No. 1.1e-38;  
 Matches 367; Conservative 0; Mismatches 136; Indels 16; Gaps 6;

QY 217 TGGAGGCCCGCAGCGCGTGGCAAGGCTGCAGCGTCCGCCCTCGAGCAGTTCA 276  
 DB 1 TGGTATCCCGCAGCGCGTGGCAAGGCTGCAGCGTCCGCCCTCGAGCAGTTCA 60

QY 277 AGATATCCGCCCGCAGCGCGTGGCAAGGCTGCAGCGTCCGCCCTCGAGCAGTTCA 336  
 DB 61 AGATATCCGCCCGCAGCGCGTGGCAAGGCTGCAGCGTCCGCCCTCGAGCAGTTCA 120

QY 337 TACACAGCAGCAGTGGAGATGCGTGAAGAACTTGTCTGTGCAGAGGAAGTGTGCCCC 396  
 DB 121 TACACAGCAGCAGTGGAGATGCGTGAAGAACTTGTCTGTGCAGAGGAAGTGTGCCCC 180

QY 397 GTGCTATTCAGAACTCTAGCCTGAACCATATTGAAGAGCTGCTGTTCTT-AGTACC 455  
 DB 181 GTGCTATTCAGAACTCTAGCCTGAACCATATTGAAGAGCTGCTGTTCTT-AGTACC 240

QY 456 TGTATAGAAATGAAATTTATGTGG---TGGCGCTATCATGGAACCGTGGTATCAGAA 512  
 DB 241 TGTATAGAAATGAAATTTATGTGG---TGGCGCTATCATGGAACCGTGGTATCAGAA 300

QY 513 GTAGTGGAGTGGATGTCGAAGAAAGTGTATTCGCCGTTCCGAGCTTAGGAGCACCTG 572  
 DB 301 AGTGTAACTGGATTTCAAGAAAGAAATGG---ATCCNCNCTTCTAACNCAAG---GACATCN 355

QY 573 TTCATCTTGGCAGCAGTGTGCCACAGCCATCTGTTGAGGTGTAGCTGGCCTTGAC 632  
 DB 356 ATCAAGNTNCTTGANATGTATNCAANAGCAATCNGTTTGAGGNACNCCGGGC---TNN 412

QY 633 TCTTTGGTCTCGTGAAGGACAAATCTTGTCTCAGGTTAACAAAGTTGTGAGGAGTGA 692  
 DB 413 ACCTTGGTCTTGTGAAGGAGNAAATCTTGNCANNTAANA----ATTCAAANAATGGG 468

QY 693 CAGAACAGTGGAGGTTGGGAAAGAACATCGATAGGATG 731  
 DB 469 CAAAAAATTGGAGGNTGGAANANANNCATTGCTNTTAAG 507

RESULT 8  
 AAF54813  
 ID AAF54813 standard; cDNA; 312 BP.  
 AC AAF54813;  
 XX AAF54813;  
 DT 15-MAY-2001 (first entry)  
 XX Nucleotide sequence of a corn Glu-tRNA reductase enzyme.  
 DE Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;  
 KW tetrapyrrolic pigment; chlorophyll; transgenic plant; ss.  
 XX Zea mays.  
 FH Key  
 FT CDS  
 FT Location/Qualifiers  
 FT 2..192  
 FT /\*tag= a  
 FT /product= "Glu-tRNA reductase"  
 FT /transl\_except= "(pos: 29..31, aa: Xaa)"  
 FT /transl\_except= "(pos: 137..140, aa: Xaa)"

FT  
 FT  
 PN  
 XX  
 PD  
 XX  
 PF  
 XX  
 PR  
 XX  
 PA  
 XX  
 PI  
 XX  
 DR  
 XX  
 DR  
 XX  
 PT  
 PT  
 PT  
 PS  
 XX  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 SQ

Sequence 312 BP; 79 A; 68 C; 90 G; 69 T; 6 other;

Query Match 9.3%; Score 179.8; DB 22; Length 312;  
 Best Local Similarity 95.0%; Pred. No. 2.3e-29;  
 Matches 207; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 224 CCAGCGCAGCGCGCTGGCAAGGCTGCCAGGCTGCCGCCCTCGAGCAGTTCAAGATATC 283  
 DB 1 CCAGCGCAGCGCGCTGGCAAGGCTGCCAGGCTGCCGCCCTCGAGCAGTTCAAGATATC 60

QY 284 CGCGCAGCGGTACATGAAGGAAAGAGTACCATAGTGTGATAGGCTTCAGTGTACACAC 343  
 DB 61 CGCGCAGCGGTACATGAAGGAAAGAGTACCATAGTGTGATAGGCTTCAGTGTACACAC 120

QY 344 AGCACCAGTGGAGATG---CGTGAAGAACTTGTCTGTGCAGAGGAAGTGTGCCCCGTCT 401  
 DB 121 AGCACCAGTGGAGATGCGGTGTAAGAACTTGTCTGTGCAGAGGAAGTGTGCCCCGTCT 180

QY 402 ATTCAGAAGAC--TCACTAGCCTGGAACCATATTGAAGAGG 438  
 DB 181 ATTCAGAAGACCTTACTAGCCTGGAACCATATTGAAGAGG 218

RESULT 9  
 ABX21149  
 ID ABX21149 standard; cDNA; 384 BP.  
 AC ABX21149;  
 XX ABX21149;  
 DT 10-FEB-2003 (first entry)  
 XX Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #3206.

KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;  
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;  
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;  
KW complex carbohydrate; gene replacement therapy; immunosuppressive;  
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;  
KW antiasthmatic; vasotropic.  
XX  
OS Homo sapiens.  
XX  
FN US2002110548-A1.  
XX  
PD 15-AUG-2002.  
XX  
PF 11-JUN-2001; 2001US-0878574.  
XX  
PR 22-NOV-1996; 96US-0753233.  
PR 03-DEC-1997; 97US-0984246.  
PR 09-SEP-1998; 98US-0149674.  
PR 14-JUN-1999; 99US-0333177.  
XX  
XX (GEMY ) GENETICS INST INC.  
XX Sullivan F, Kriz R, Kumar R;  
XX WPI; 2003-066673/06.  
XX  
PT New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)  
PT peptide, for manufacturing complex carbohydrates, or as targets for  
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant  
PT rejection -  
XX  
PS Disclosure; SEQ ID NO 3208; 6pp; English.  
XX  
XX The invention relates to a composition comprising a human GDP-mannose  
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying  
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation  
CC in a mammalian subject and for treating or ameliorating diseases affected  
CC by the level of cellular fucosylation or diseases affected by the  
CC fucosylation of glycoconjugates. These diseases include arthritis,  
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or  
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also  
CC useful for manufacturing complex carbohydrates and as targets for  
CC screening small molecule antagonists of the activity of the enzyme. The  
CC polynucleotide is useful in developing an assay for defects in the  
CC enzyme, as well as in gene replacement therapy. Sequences  
CC ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding  
CC human GM4,6D peptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html.  
XX  
XX Sequence 384 BP; 128 A; 61 C; 95 G; 100 T; 0 other;  
Query Match 6.5%; Score 125.6; DB 25; Length 384;  
Best Local Similarity 64.6%; Pred. No. 1.3e-17;  
Matches 204; Conservative 0; Mismatches 109; Indels 3; Gaps 1;  
QY 1366 AGAAGCTGAGGTCGTACCGGACAGCATCGAGGCTCGAGAGTCCCTGCGACA 1425  
DB 1 AGAATTGAGGCTTATGCTGAAAGAAATCAGGCTTGTGAGCTTGAAGAGTCTTAGGTA 60  
QY 1426 AAGTAGGTGAGGACGCGCTTCCACCAAGAGATGAGGAGAGCCATCGAGGAGCTGAGCAGC 1485  
DB 61 AGATGGTGATGATATAC---CAAGAAACCGGAGAGCTGTGATGACCTTAGTCGG 117  
QY 1486 GCATGCTTAAAGCTCTCCATGCGCCCTCGAGACCTGAGGTGCGACGCGCAGACA 1545  
DB 118 GTATAGTGAATAAGTTGCTTCAATGTGTCCAAATGCAACATTTAAGGTGTGATGGAACGACA 177  
QY 1546 GCGGACCCCTTGCAGAGAGCTTCGAGACATCGACGCTTCAACCGGATGTTTCAGCCTCG 1605  
DB 178 GCCGGACTTCTAGTGAGACACTGGAGAACATGAATGCTTTGAATAGGATGTTCAACCTTG 237

QY 1606 ACATGGAGAAGCGCATCATCGAGCAGAGATCAAGGCCAAGTGTGAGAGACACAAAAT 1665  
DB 238 AGACAGAAATATCTGTTTTGGAGGAGAAAGATTCGAGCAAGAGTTCGAACAAAACCGAAAT 297  
QY 1666 GAGGCCAGGAAGCAAT 1681  
DB 298 GAAATCTTAACCAAT 313  
RESULT 10  
ABX23406  
ID ABX23406 standard; cDNA; 275 BP.  
XX  
AC ABX23406;  
XX  
DT 10-FEB-2003 (first entry)  
XX  
DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #5463.  
XX  
XX Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;  
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;  
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;  
KW complex carbohydrate; gene replacement therapy; immunosuppressive;  
KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;  
KW antiasthmatic; vasotropic.  
XX  
OS Homo sapiens.  
XX  
PN US2002110548-A1.  
XX  
PD 15-AUG-2002.  
XX  
PF 11-JUN-2001; 2001US-0878574.  
XX  
PR 22-NOV-1996; 96US-0753233.  
PR 03-DEC-1997; 97US-0984246.  
PR 09-SEP-1998; 98US-0149674.  
PR 14-JUN-1999; 99US-0333177.  
XX  
XX (GEMY ) GENETICS INST INC.  
XX Sullivan F, Kriz R, Kumar R;  
XX WPI; 2003-066673/06.  
XX  
PT New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)  
PT peptide, for manufacturing complex carbohydrates, or as targets for  
PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant  
PT rejection -  
XX  
PS Disclosure; SEQ ID NO 5465; 6pp; English.  
XX  
XX The invention relates to a composition comprising a human GDP-mannose  
CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying  
CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation  
CC in a mammalian subject and for treating or ameliorating diseases affected  
CC by the level of cellular fucosylation or diseases affected by the  
CC fucosylation of glycoconjugates. These diseases include arthritis,  
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or  
CC infection. The GM4,6D peptide or a polynucleotide encoding it is also  
CC useful for manufacturing complex carbohydrates and as targets for  
CC screening small molecule antagonists of the activity of the enzyme. The  
CC polynucleotide is useful in developing an assay for defects in the  
CC enzyme, as well as in gene replacement therapy. Sequences  
CC ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding  
CC human GM4,6D peptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 275 BP; 60 A; 51 C; 84 G; 80 T; 0 other;  
Query Match 6.4%; Score 122.6; DB 25; Length 275;



	Best Local Similarity	71.4%; Pred. No. 5.3e-17;	Matches 175; Conservative 0; Mismatches 69; Indels 1; Gaps 1;
Qy	603	CATCTGTTTCAGGTGTCAGCTGGCCTTGACTCTTGGTTCTCGGTGAAGGACAAATCCTT	662
Db	7	CAICTATTGAAGTGGCGTCAGGCCTTGACTCACTGTTCTTGGGAAGGTCAAATTCCTT	66
Qy	663	GCTCAGGTTAAAACAAGTTGTGAGGAGTGGACAGAACAGTGGAGCTTTGGGAAGAACAATC	722
Db	67	GCTCAGGTGAAGCAGGTTGTGAAGCTGGAACAGGAGTGCCTGGTTTTGATAAGAAATC	126
Qy	723	GATAGGATGTTCAAGGATGCNAATCACCTCTGGAAGCGTGTCCGACGGCGACACCACATA	782
Db	127	AGTGGTTTGTTCAAACAGCGCGATATCGTTGGGAAGCGGGTTAGAACCGAGACTTAACATT	186
Qy	783	TCATCTCGTGCTGTTTCTGTTCAGTTCAGCGCGGTGAACTGGGCCCTGTGATCAAGCTCCG	842
Db	187	TCATCTGGATCAGTTTCTGTAAGCTTCGGCTGCTGTGGAGCT-GCATGTATGAGTACCG	245
Qy	843	AAGTC 847	
Db	246	GATTC 250	

RESULT 11	
AAQ88151	
ID	AAQ88151 standard; DNA; 1549 BP.
XX	
XX	AAQ88151;
XX	
XX	02-NOV-1995 (first entry)
XX	
XX	Xanthomonas L-glutamyl tRNA reductase.
XX	
XX	Xanthomonas campestris; L-glutamyl tRNA reductase; vitamin B12;
KW	glutamic acid-1-semialdehyde; 5-aminolevullinate (ALA); haem;
KW	chlorophyll; phycobillin; recombinat production; ds.
XX	
XX	Xanthomonas campestris.
OS	
XX	
XX	Key Location/Qualifiers
FH	212..1492
FT	/*tag= a
FT	
XX	
PN	JP07031480-A.
XX	
XX	
PD	03-FEB-1995.
XX	
XX	27-JUL-1993; 93JP-0184709.
PF	
XX	
XX	27-JUL-1993; 93JP-0184709.
PR	
XX	
XX	(COSM-) COSMO SOGO KENKYUSHO KK.
PA	
XX	
XX	WPI; 1995-109535/15.
DR	P-PSDB; AAR74619.
DR	
XX	
XX	DNA fragment encoding L-glutamyl tRNA reductase - derived from
PT	Xanthomonas sp., useful for production of 5-amino:levullinate,
PT	vitamin-B12, chlorophyll, etc.
PT	
XX	
XX	Claim 3; Pages 7-8; 9pp; Japanese.
PS	
PS	
XX	
XX	AAQ88151 encodes AAR74619 Xanthomonas campestris L-glutamyl tRNA
CC	reductase. The DNA as part of an expression vector can be used to
CC	transform a host cell, for the recombinant prodn. of the protein.
CC	The protein can be used in the prodn. of vitamin B12, glutamic
CC	acid-1-semialdehyde, 5-aminolevullinate, haem, chlorophyll and
CC	phycobillin.
CC	
XX	
SQ	Sequence 1549 BP; 291 A; 516 C; 514 G; 228 T; 0 other;

Query Match 5.2%; Score 99.4; DB 16; Length 1549;

Best Local Similarity		46.2%	Pred. No. le-11;		
Matches	Conservative	0;	Mismatches	561;	Indels
					45; Gaps
Qy	300	AAGGAAAGGAGTACCATAGCTGTGATAGGCTCAGTGTACACAGACACAGTGGAGATG	359		
Db	203	AATGGAACAATGACTGTTGGTGTCTCGAAGTATCACCAGACCGACCTGTGGACTG	262		
Qy	360	CGTGAAGAACTGTGTTGAGAGGAACTGTGGCCCGCTGTATTCAAGAACTCACTAGC	419		
Db	263	CGGAAACGCGCGGCTTCGAGGTGATGCGCTGCGCGCGCTCGAATCGCTCGTGCG	322		
Qy	420	CTGAACCATATTGAAGAGGCTGCTGTTCTTAGTACCTGTATAGAAATGGAATTTATGTG	479		
Db	323	CTGCCCAGGTAAAGCAGGCGCGGCTGCTGTCCACCTGCAACCGTACCGAGTTGTATGCG	382		
Qy	480	GTGCGCTATCATGTGAACCGTGTGTATCAGAGAAAGTAGTGACTGTGATGTCGAAGAAAAGT	539		
Db	383	ATGGCC-----GAGAGCGCACAGCTGGTCACTCGGCTG	418		
Qy	540	GGTATTCCTCCGCTTCGAGCTTAGGGAGCACTGTTTCACTTTGGAAGCAGTGATGCCACA	599		
Db	419	GAAGCCACACGCGCGCTCTGAGCGGTTATCTGTACCAGCACAGGAAGCCGAGCGGTG	478		
Qy	600	CGCCATCTGTTGAGGTGTGAGTGGCTTTGACTCTTTTGGTTCTCGGTGAAGGACAAATC	659		
Db	479	CGTCATCTGTTCCGTGTGCGCACCGGCTGGATTTCGATGTTATGGAGAAACCCAGATC	538		
Qy	660	CTTGCTCAGGTTAAACAAAGTTGTGAGAGTGGAGACAGACAGTGGAGCTTTGGGAAGAAC	719		
Db	539	CTCGGCAGGTGAAGGACCGCTCGGCAGTGGCGCGGCGCACGGCACGCTGGGCGCGGG	598		
Qy	720	ATCGATAGGATGTTCAAGGATGCAATCACCTGCTGGAAGGCTGTGCGACGAGACCAAC	779		
Db	599	TTGGAACGGCTGTTCCAAAGACCTTTTCGGTGGCCAAAGCGCGGTACCGACACCCGT	658		
Qy	780	ATATCATCTGGTCTGTTTCTGTGAGTTCAGCGCGGTGAACTGGGCCCTGATGAAGCTT	839		
Db	659	GTGCGGCCAAATCCGGTGTCCGTCGCATCCACGCGGTGCGGCTGGCACAAAGACTCGTTT	718		
Qy	840	CCGAGTCTGAAGCATGTGAGTAGATGCTTCTGATTTGGTGTGTGTAATATGGGAAG	899		
Db	719	GCCCGGCTCAACGAATCGACGG-----TGTTGCTGATCGTGGCGGAAACCATCGAA	772		
Qy	900	CTAGTGATCAAAATCTGGTTGCAAAAGGATCAAGAAGTGTGTTGGTGAACCGCTCC	959		
Db	773	CTGCGGCCAAGCATCTGAGCGAAGCGCGGTGCGCGGCTGCTGATCGCAACCGCACC	832		
Qy	960	GTGGAAGGGTGGATGCTATTCTGTGAGGAGATGAAAGATATAGAGATCGTGTACAGGCCT	1019		
Db	833	CTGGCCACGCGCAGAC-----GCTCGCCAGCCAGCATCGCGGCTTCGCTTGGCG	883		
Qy	1020	CTCTCAGACATGATCAAGCTGTCTGAGCTGTGTGAGTGTGTGTTCCACGACCGCATCT	1079		
Db	884	CTGACCGATCTGGAACGCCACCTGGCGCGGCGGATGTGGTGTCTCGGCCACCGCTGCA	943		
Qy	1080	GAAACTTCAATTGTTCCGAAAGAAACACGAGAGGCACTCCCCCTGCTCTGTGATCTATG	1139		
Db	944	CGTAGCCGCTGGTGAC-----CCGCGCACAGGTGGAAACAGGCATTTGCGCGCACGCAAG	997		
Qy	1140	GGAGGTGTTCCGCTGTTTGTGCACATATCTGTCTCCCGAGGAATGTGACGCGCATGTGTGCT	1199		
Db	998	CGCAACCCGATGCTGTGTTTCGACTCTGGCGGTGCGCGCGATATCGAGGCTCGGTGGCG	1057		
Qy	1200	GAATGTTGGCGTGCAGAGTGTAACAATGTGACGACTTTGAAAGAGGTGTGGGAAGCAAC	1259		
Db	1058	GAATTGAGCGACGCCCTACCTGTACACGGTGGACGACCTGGAAACGCGCGTTCGAAGACAAC	1117		
Qy	1260	AAGGAGGACGGCTCAGGAAAGCAATGGAAGGCGCAGACAAATCATACCGAAGAACTGAGA	1319		
Db	1118	CGCCGTGGCGCGCGAAGCGGCCGACCGAAGCCATCATTCGACCTGCAAGTGGCG	1177		
Qy	1320	CGGTTCCAGGCAATGGAGGACTCGCTGGAGACCGTTCCGACCATCAAGAGCTGAGGTGCG	1379		

Db 1178 CGCTATGTGAAACCTTGAGGCCAACGCGGCCAGGACCGCTCAAGCGGCTGCGGCC 1237  
QY 1380 TACGCGGACAGATCAGGCGCTCGAGCTCGAGAGTGCCTGCAGAA 1426  
Db 1238 TTGGCGATAGCACCGCGACGAGCTGTGGCCAAAGCGCGCGCAGCA 1284

## RESULT 12

ID ABK73683  
XX AC ABK73683  
XX DT 13-AUG-2002 (first entry)  
XX DE Bacillus licheniformis genomic sequence tag (GST) #974.

XX KW Differential gene expression; genomic sequenced tag; GST;  
XX KW altered culture condition; environmental stress;  
XX KW physiological provocation; ds.  
XX OS Bacillus licheniformis.  
XX PN WO200229113-A2.  
XX PD 11-APR-2002.  
XX PF 05-OCT-2001; 2001WO-US31437.  
XX PR 06-OCT-2001; 2000US-0680598.  
XX PR 27-MAR-2001; 2001US-279526P.

XX PA (NOVO) NOVOZYMES BIOTECH INC.  
XX PA (NOVO) NOVOZYMES AS.  
XX PI Berka R, Clausen IG;  
XX DR WPI; 2002-416684/44.  
XX PT Monitoring differential expression of several genes in first Bacillus  
cell relative to expression of same genes in one or more second  
Bacillus cells, by using substrate containing Bacillus genomic  
sequenced tag array -  
XX PS Claim 4; SEQ ID NO 974; 200pp; English.

CC The invention describes a method of monitoring differential expression of  
genes in a first Bacillus cell relative to expression of the genes in  
other Bacillus cells, comprising hybridising labelled nucleic acid probes  
isolated from Bacillus cells to a substrate containing array of Bacillus  
genomic sequenced tags (GST), examining the array, and determining  
relative gene expression by an observed hybridisation reporter signal of  
a spot in the array. The method is useful for measuring the expression of  
genes in a first Bacillus cell relative to expression of the same genes  
in one or more second Bacillus cells. The method is useful for monitoring  
global expression of several genes from a Bacillus cell, discovering new  
genes, identifying possible functions of unknown open reading frames and  
monitoring gene copy number variation and stability. Monitoring changes  
in expression of genes may be used to provide a representation of the way  
in which Bacillus cells adapt to changes in culture conditions,  
environmental stress or other physiological provocation. Extensive  
follow-up characterisation is unnecessary, when one spot on an array  
equals one gene or one open reading frame, since sequence information is  
available. This sequence represents a genomic sequence tag (GST) used in  
the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at  
ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1360 BP; 367 A; 293 C; 374 G; 326 T; 0 other;

## Query Match

4.5%; Score 86.6; DB 24; Length 1360;

Best Local Similarity 44.7%; Pred. No. 5.7e-09;  
Matches 485; Conservative 0; Mismatches 579; Indels 21; Gaps 3;  
QY 315 ATAGCTGTGATAGGCTCAGTGTATACACAGCACCAGTGGAGATCGGTGAAAAAATTGCT 374  
Db 7 ATACTTGTCTGTTGGATTGATTACAAATCAGCCCTGTTGATATACGTGAAAATTAACT 66  
QY 375 GTTGCAGAGAACTGTGCCCCCGCTATATCAAGAACTCACTAGGCTGAACCATATTTGAA 434  
Db 67 TTTTCAGCGATGAGCTGGCGCAGCGATGCGCAATTGAAAGAAAGAAAAAGCACTACTT 126  
QY 435 GAGGCTGCTGTTCTTAGTACCTGTAATAGAAATGGAATTTATGTGGTGGCGCTATCATGG 494  
Db 127 GAGAAATCATCTGCTGACATGCAACCGGACAGAGCTGTATGCAAGTCTGTTGATCAGCTG 186  
QY 495 AACCGTGGTATCAGAGAAAGTAGTGGACTGTGCAAGAAAGAGTGGTATTCCCGCTTCC 554  
Db 187 CACACCGGACGTTATTATGAAATGTTTTTGGCGAATCGTTCGGATTAAAGCAAGAG 246  
QY 555 GAGCTTAGGAGACCTGTTTCATCTTGGCAAGCAGTGTATGCCACACGCCATCTGTTTGGAG 614  
Db 247 GACATTTTCGCTTACTTTGAAGTTTTATGAAAACGACGGGGCGCTTGAACACACCTGTTCCGC 306  
QY 615 GTGTCAGCTGGCCTTGACTCTTTCGTTCTCGGTGAGGACAAATCCTTGTCTCAGCTTAA 674  
Db 307 GTATCGTGGATTGGATTGATGTTTATGAGAAACGACAGATTTCGGGCAAGTGGCT 366  
QY 675 CAAGTTGTGAGGAGTGGACAGAACAGTGGAGGCTTTGGGAAAGAAACATCGATAGGATGTC 734  
Db 367 ACAAGTTTAAAGCTGGCTCAGGAGAGAGAAATGTTGTAACGTTTCACTATTGTTT 426  
QY 735 AAGGATGCAATCACTGTGGAAGCGGTGTCGCGAGCGAGACCAATATCATCTGTTGGTGT 794  
Db 427 AAGCAGGCGGTCAAGCTGCGAAAGCGCTGCCATGCCGAAACGACATTTCTCAAAATGCT 486  
QY 795 GTTCTGTCACTTCAGCGCGGTTGAACTGCGCCCTGATGAGCTTCCGAGCTTCCGAGTCTCAAGCA 854  
Db 487 GTATCCGTCAGCTATGCCGCGCTGAGCTTCCAGGAAATCTTCGGCGACCTTTCCCGAT 546  
QY 855 CTGTCACTAGGATGCTTCTGATTGGTGTCTGTTAAAATGGGAAAGCTAGTATCAAAACAT 914  
Db 547 AAACACG-----TGCTGATTTCTCGGGCGGCAAAATGGGTGAGCTTCGGTTTCAAAAC 600  
QY 915 CTGTTGCCAAAGGATGCAAGAAAGTTGTTGTGTGAACCGCTCGTGAAGAGGTGTAT 974  
Db 601 CTGACGCGCCACGGAATCGGACAGGTTACGCTGATCAACAGAACATTTTCAAAAGCGAAG 660  
QY 975 GCTATTTCGTGAGGAGTGAAGATATAGAGATCGTGTACAGGCCTCTCTCAGACATGTAT 1034  
Db 661 GAGCTTCCGCGCGCTTTTCAGCGCGCGGAAAGCTTGAA-----TGAGCTGCAA 711  
QY 1035 CAAGCTGTCTGTAAGCTGTATGCTGTTTACCAGCACCGCATCTGAAACTTTCATTTGTC 1094  
Db 712 TGCCTTTGATGGAAGCGGATTTCTCATCACTGTCAGCCGCGCGCAGCGGTATGTCGTC 771  
QY 1095 GCAAAAGAACACGAGGAGGCACTCCCGCTGCTCTGATATATATGAGGAGTGTTCGCGCTG 1154  
Db 772 ACAAAGAGATGATCGAAACATGTCAATAAGCT-----TCGAAAGGCTGTCCGCTTTT 825  
QY 1155 TTTCTGCACATATCTGTCGCCAGAAAGTTCAGCGCATGTGTCTGAGTTGGCGCTGCA 1214  
Db 826 ATGTTGACATATGCGTGGCGCGGATCTTGACCCGGCTCTCAGCGAAGTGGAGCGGCTC 885  
QY 1215 CGAGTGTACATGTGACGACCTTGAAGAGAGTGTGGAAGCAACAGAGGACCGGCTC 1274  
Db 886 TTCCTGTACGATATTGACGATCTTGAAGGATCTGCGAAGAAATTTGAAGAGACCGCAG 945  
QY 1275 AGGAAAGCAATGAGGCGGAGACAAATCATCCGAAAGACTGAGACCGGTTCAGGCAATGG 1334  
Db 946 GCCGTTGCCGAGAGGTGGAGCTGATTATTGAAGCTGAGATCGTTTCTTTTAAACAATGG 1005  
QY 1335 AGGAGCTCGCTGGAGACCGTTCCGACCATCAAGAGCTGAGTGTGTACGGGACGAGATC 1394

Db 1006 CTGAATACGCTCGCGCTCGTCCGCTCATCTCGGCTCTCAGGAAAGCCCTTGACGATC 1065  
 Qy 1395 AGGCG 1399  
 Db 1066 CAGGC 1070

## RESULT 13

ABK78210  
 ID ABK78210 standard; DNA; 1131 BP.

XX AC ABK78210;

XX DT 13-AUG-2002 (first entry)

XX DE Bacillus clausii genomic sequence tag (GST) #1053.

XX XX Differential gene expression; genomic sequenced tag; GST;

KW altered culture condition; environmental stress;

KW physiological provocation; ds.

XX OS Bacillus clausii.

XX XX WO200229113-A2.

XX XX 11-APR-2002.

XX PF 05-OCT-2001; 2001WO-US31437.

XX XX 06-OCT-2000; 2000US-0680598.

PR 27-MAR-2001; 2001US-279526P.

XX XX (NOVO ) NOVOZYMES BIOTECH INC.

PA (NOVO ) NOVOZYMES AS.

XX FI Berka R, Clausen IG;

XX XX WPI; 2002-416684/44.

XX XX Monitoring differential expression of several genes in first Bacillus  
 cell relative to expression of same genes in one or more second  
 PT Bacillus cells, by using substrate containing Bacillus genomic  
 PT sequenced tag array -

XX PS Claim 11; SEQ ID NO 5501; 200pp; English.

XX CC The invention describes a method of monitoring differential expression of  
 CC genes in a first Bacillus cell relative to expression of the genes in  
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
 CC isolated from Bacillus cells to a substrate containing array of Bacillus  
 CC genomic sequenced tags (GST), examining the array, and determining  
 CC relative gene expression by an observed hybridisation reporter signal of  
 CC a spot in the array. The method is useful for measuring the expression of  
 CC genes in a first Bacillus cell relative to expression of the same genes  
 CC in one or more second Bacillus cells. The method is useful for monitoring  
 CC global expression of several genes from a Bacillus cell, discovering new  
 CC genes, identifying possible functions of unknown open reading frames and  
 CC monitoring gene copy number variation and stability. Monitoring changes  
 CC in expression of genes may be used to provide a representation of the way  
 CC in which Bacillus cells adapt to changes in culture conditions,  
 CC environmental stress or other physiological provocation. Extensive  
 CC follow-up characterisation is unnecessary, when one spot on an array  
 CC equals one gene or one open reading frame, since sequence information is  
 CC available. This sequence represents a genomic sequence tag (GST) used in  
 CC the method of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1131 BP; 344 A; 225 C; 284 G; 278 T; 0 other;

Query Match 4.3%; Score 83; DB 24; Length 1131;

Best Local Similarity 46.9%; Pred. No. 3.2e-08;  
 Matches 380; Conservative 0; Mismatches 410; Indels 21; Gaps 3;

Qy 555 GAGCTTAGGGAGCACCTGTTTCATCTTGCAGACGAGTGTATGCACACGCACTCTGTTTGAG 614  
 Db 47 GACGTAAACGCCGTATTTGGTTATCCGTGAAAAATGACCATGCAATGAAACACCTTTTCGC 106

Qy 615 GTGTCAGCTGGCCTTGACTCTTTGGTTCTCGGTGAGGACAAATCCCTTCTCAGCTTAAA 674  
 Db 107 GTTCTTGTGTTTAGACTCTATGGTATCGCGGAAACGCAATTTCTTTGGACAAGTGAAG 166

Qy 675 CAAGTTGTGAGGAGTGGACACAGACAGTGGAGGCTTTGGGAAAGAAACATCATAGGATGTTTC 734  
 Db 167 CAGGCATTTTGTCTTGGCCAAAGAGAAAGTCAACAGCACTGTGTTTAAACAGCTATTT 226

Qy 735 AAGGATGCAATCACTGCTGGAAAGCGTGTCCGCGAGCGAGACCAATATCATCTCGGTGCT 794  
 Db 227 AAACAAGCGGTTACGCTCGGCAACGTTGTGCACTCGTCAACGGAATTTCTTCACAAGCC 286

Qy 795 GTTTCTGTCACTTCAGCGCGGTTGAACTGCGCCCTGTATGAGCTTCCGAAGTCTCGAAGCA 854  
 Db 287 GTTCTGTATGCTATGACGGCTGAGAACTGGGAAAGAAAAATTTTCGGCACGTTTTTCAGCA 346

Qy 855 CTGTCAGCTAGGATGCTTCTGATTGGTGTCTGTTAAAATGGGAAAGCTAGTGATCAAAAT 914  
 Db 347 AAACATG-----TGCTTATTTTAGGCGCTGGAATAATGAGGAACTGACGCCAAACAT 400

Qy 915 CTGTTTCCAAAGGATGCAAGAGTTGTTGTGTGTAACCGCTCGTGGAAAGGTTGAT 974  
 Db 401 CTTTATTTGGAATGGGCTGCTTCCATTAGTGTCTATGAACCGGACAAAGAAAGAAACCGGTG 460

Qy 975 GCTATTCGTGAGGAGATGAAAGATATAGAGATCGTGTACAGGCTCTCTCAGACATGAT 1034  
 Db 461 GAACCTGGAAAGCCAGTTTCTGGAAACGGCT-----CGGTCTTTTGTCTGAATTTGAAT 511

Qy 1035 CAAGCTGCTGCTGAAAGCTGATGTCGTTTCAACAGCACCGCATCTGAAACTTCATTGTTTC 1094  
 Db 512 GATGCTTTTAAAGAAAGCAGATGTTGTGATTAGCTTCCACGGGGCCCGGCACTTTGTTGTC 571

Qy 1095 GCAAAAGAACACGAGGCACTCCCGCTGCTCTCTGATATCTATGGGAGGTTGTTCCGCTG 1154  
 Db 572 AAAAAGGAAATGCCGACGCTGCC-----TTAAAAAAGGAAAGGACGCGCGCTGTT 625

Qy 1155 TTTGTCGACATATCTGTCGCCGGAATGTACGCGCATGTGTCTGAAGTTGGCGCTCA 1214  
 Db 626 GCGATTGATATCGCAGTGCCACGGGATATTGACCCCGGAAATTCACAAATTTAGGACGTT 685

Qy 1215 CGAGTGTACAAATGTGACGACGACTTGAAAGAGGTGGTGAAGCAACAAAGGAGGACCGGCTC 1274  
 Db 686 TACTTATACGACATTGATGACTTACAGAACATTTGTTGAAGCCATATAAAGAGCGCAAA 745

Qy 1275 AGGAAAGCAATGGAGCGCGACAAATCATCAACGAGAACTGAGACGGTTCCAGGCGATGG 1334  
 Db 746 AAAGAGGCTGAAAAAATCGGTGTTATGATTGAGGACAGATCGATGACTTTTAAAGCATGG 805

Qy 1335 AGGACCTCGTGGAGACCGTTCCGACCATCA 1365  
 Db 806 CTAACACACTAGGGGTGGTGCCTCTTTATTA 836

## RESULT 14

AAZ54023

ID AAZ54023 standard; DNA; 1248 BP.

XX AC AAZ54023;

XX XX 21-MAR-2000 (first entry)

XX DE Neisseria meningitidis ORF 622 partial DNA sequence SEQ ID NO:1995.

XX XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy; ds.

XX OS *Neisseria meningitidis*.  
 XX PN W09957280-A2.  
 XX PD 11-NOV-1999.  
 XX PF 30-APR-1999; 99WO-US09346.  
 XX PR 01-MAY-1998; 98US-0083758.  
 XX PR 31-JUL-1998; 98US-0094869.  
 XX PR 02-SEP-1998; 98US-0098994.  
 XX PR 02-SEP-1998; 98US-0099062.  
 XX PR 09-OCT-1998; 98US-0103749.  
 XX PR 09-OCT-1998; 98US-0103794.  
 XX PR 09-OCT-1998; 98US-0103796.  
 XX PR 25-FEB-1999; 99US-0121528.  
 XX PA (CHIR) CHIRON CORP.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;  
 PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX DR WPI; 2000-062150/05.  
 XX DR P-PSDB; AAY75261.  
 XX PT Novel *Neisserial* polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics -  
 XX PS Claim 7; Page 997; 1453pp; English.  
 XX CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC *Neisserial* bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX SQ Sequence 1248 BP; 300 A; 305 C; 363 G; 280 T; 0 other;

Query Match 4.1%; Score 78.6; DB 21; Length 1248;  
 Best Local Similarity 47.5%; Pred. No. 3e-07;  
 Matches 317; Conservative 0; Mismatches 329; Indels 21; Gaps 2;

QY 325 TAGCCCTCAGTGATACACAGCAGCAGTCGATGAGATGCGTGAATAAATCTGCTGTCAGAGG 384  
 DB 17 TCGGACTCAATCATCAAAACCCGACCTTTAAGCATACCGGAAAGCTGGCGTTTGGCCGG 76  
 QY 385 AACTGTGGCCCGGTGCTATTCAAGAACTCACTAGCCTGAACCATATTGAAGAGGCTGCTG 444  
 DB 77 CCTGCTGCCCGAAGCCGTCGGCAATCTTGCCGAAGCAATGCGGCAACGAGGCGGTAA 136  
 QY 445 TTCTTAGTACTGTAATAGAAATGAAATTTATGTGTGGCGCTATCATGAAACCGTGGTA 504  
 DB 137 TCCTTTCTAGCTGCAACCGTACCGAGCTTTACTGTGTAG-----GTGATT 181  
 QY 505 TCAGAGAGTAGTGCATGATGTCGAAGAAAGTGTATTCCGGCTTCGAGCTTAGGG 564  
 DB 182 CGGAAGAAATCATCCGTTGGCTGCGACATATCAAGCCCTTCCCATAGAGAATCAGCC 241  
 QY 565 AGCACCTGTTTCATCTTCGGAAGCAGTGTATGTCACACCGCCATCTGTTTGGTGTGACCTG 624  
 DB 242 CCTACCTTTATCTTTGGGATGCGAGGACTGTGCGCCATGCTTTCCGCGTCCCTGCG 301

QY 625 GCCTTGACTCTTTGGTTCTCGTGAAGGACAAATCCTTGCTCAGGTTAAACAAGTTGTGA 684  
 DB 302 GCTTGGATTTCGATGTTGGGCCAGCGCGCAGATTTTAGGACAGATTAAAGGATCGGTCA 361  
 QY 685 GGAGTGGACAGAAACAGTGGAGGCTTGGGAAGAAGACATCGATAGGATGTTCAAGGATGCAA 744  
 DB 362 GGGTTGCTCAAGACGAGAAAGTATGGTAAGAACTCAATGCCCTGTTCCAAAAAACCCT 421  
 QY 745 TCACGTCTGGAAGCGGTGTCGACGAGACCAACATATCATCTGGTGTCTGTTCTGTCTGA 804  
 DB 422 TTTCTGTTGCTAAAGAGTCCGTACCGATACTGCGCGGCGGAAAACTCGGTTTCCATGG 481  
 QY 805 GTTCAGCGCGGTTGAACCTGCCCCCTGATGAAGCTTCCGAAGCTTGAAGCACTGTGACGTA 864  
 DB 482 CTTCCGCTTCGCTCAAGTTGGCAGAGCAGATTTTCCCGACATCGGCGATTTGAATG--- 538  
 QY 865 GGATGCTTCTGATTGGTGTGTAATAATGGAAAGCTAGTATCAAAACATCTGGTTGCCA 924  
 DB 539 ---TCTTGTATTATCGGTCCGGTGAGATGATTGAGCTGGTTGCCACTTATTTTCCGCCA 595  
 QY 925 AAGATGCAAGAGGTTGTTGTGTGAACCGCTCCGTTGGAAAGGGTGGATGCTATTCTGTG 984  
 DB 596 AAAGTCCCGGCTGATGACGTTTGCCAAACCGAGCTGGCGCTGCACAGGAGTTGTGCG 655  
 QY 985 AGGAGAT 991  
 DB 656 ACAAGCT 662

RESULT 15  
 ABQ67195  
 ID ABQ67195 standard; DNA; 495269 BP.  
 XX AC ABQ67195;  
 XX DT 29-AUG-2002 (first entry)  
 XX DE *Listeria innocua* contig DNA sequence #8.  
 XX KW Antibacterial; *Listeria*; food contamination; mutational analysis;  
 XX KW infection; da.  
 XX OS *Listeria innocua*.  
 XX PN W0200228891-A2.  
 XX PD 11-APR-2002.  
 XX PF 04-OCT-2001; 2001WO-FR03061.  
 XX PR 04-OCT-2000; 2000FR-0012697.  
 XX PA (INSP) INST PASTEUR.  
 XX PA (CNRS) CNRS CENT NAT RECH SCI.  
 XX PI Kunst F, Glaser P;  
 XX DR WPI; 2002-332479/37.  
 XX PT New genomic sequences from *Listeria* species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators -  
 XX PS Claim 5; SEQ ID 8; 180pp; French.  
 XX CC The present invention relates to nucleic acid sequences  
 CC (ABQ67188-ABQ671212) from *Listeria* sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of *Listeria* (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
 CC treating infections by *Listeria*, and are useful as immunogens in

CC anti-listeria vaccines.	
CC Note: The sequence data for this patent did not form part	
CC of the printed specification, but was obtained in electronic format	
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX	
SQ	Sequence 495269 BP; 159529 A; 88949 C; 93135 G; 153754 T; 2 other;
Query Match 3.9%; Score 75.4; DB 24; Length 495269;	
Best Local Similarity 46.1%; Pred. No. 1.1e-05;	
Matches 293; Conservative 0; Mismatches 336; Indels 6; Gaps 1;	
Qy	324 ATAGGCCTCAGTGTACACAGCAGCAGTGGAGATGCGTGAAGAACTTGTGTTGCAGAG 383
Db	160090 ATGGGGCTGAATCATCACAGCAGCAGTGGAGATGCGTGAAGAACTTGTGTTGCAGAG 160149
Qy	384 GAACCTGTGGCCCGTGTATTCAAGAACTCACTAGCCTGAACCATATTGAAGAGCGCTCT 443
Db	160150 ACGGAAGAGAAATGGCCTTAGTAACTTACTGAGGAAAGATATCCTCGAAATGTT 160209
Qy	444 GTTCTTAGTACTCTGAATAGAAATTTATGTGTGGCGCTATCATGGAACCGTGGT 503
Db	160210 ATTATTCGACATCAATCGTACGAGATTGTGGCAGTTGTGATCAATACATACAGGC 160269
Qy	504 ATCAGAGAGTAGTGGAGCTGGATGTGGAAGAAAGTGGTATTCGCGTTCGAGCTTAGG 563
Db	160270 AGATATTACTTTAAACCGCTTTATGGCCAAATTGGTTTCAATGGATATGGAAGAAATCGAG 160329
Qy	564 GAGCACCTGTTTCATCTTCGGAAGCAGTGATGCCACACGCCATCTGTTGAGGTGTCACT 623
Db	160330 CCTTATTGTTTTCATGAAGAGCAGAGCGCTCAATCATTTATATAAGTAACCTGCG 160389
Qy	624 GGCCTTGACTCTTTGGTTCTCGTGAAGGACAAATCCTTGCTCAGGTTAAACAAGTTGTG 683
Db	160390 GGACTCGATTCACTTGTACTCGGAGAAACGCAATTCCTCGACAAAGTAACATGCATT 160449
Qy	684 AGGAGTGACAGAACAGTGGAGGCTTGGGAAGAACATCGATAGGATGTTCAAGGATGCA 743
Db	160450 GAAATCGGAAACAGACTGGGAACAACAGGCACACTCTTAAATAAACTTTTCGGGAAGTA 160509
Qy	744 ATCACTGCTGGAAGCGTGTCCGAGCAGAGACCAACATATCATCTGGTGTGTTCTGTC 803
Db	160510 GTTACTTTTGCAGAAAGTGATCATCATACGAAATTAACGAAATGCGGTTTCGGTC 160569
Qy	804 AGTTCAGCGCGGTTGAACCTGGCCCTGATGAAGCTTCCGAAGTCTGAAGCACTGTCACT 863
Db	160570 AGTTATCAGCTGTAGAAGTTGCC-----AABAACTATATGCTTATTAGATAATAA 160623
Qy	864 AGGATGCTTCTGATGTGTGCTGTAAATGGGAAAGCTAGTATCAAAACATCTGGTTGCC 923
Db	160624 AAAATCGTCTCTTATTGGAGCAGGGGAAATGAGCGAACTTGCTTTACAAATCTTGCTGGA 160683
Qy	924 AAAGGATGCAAGAGGTTGTTGGTGAACCGCTC 958
Db	160684 AGTGGCATTGAGATATTACTATCATCAATCGCAC 160718

Search completed: December 15, 2003, 12:57:32  
Job time : 553 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 12:57:36 ; Search time 1512 Seconds  
(without alignments)  
4229.236 Million cell updates/sec

Title: US-10-018-902-3

Perfect score: 1924

Sequence: 1 ccacgcgtccgcatcaataa.....aaaaaaaaaaaaaaaaaag 1924

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US05\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	679.6	35.3	1632	10	US-09-938-842A-1201
2	125.6	6.5	384	10	US-09-878-574-3208
3	122.6	6.4	275	10	US-09-878-574-5465
4	96.6	4.5	1360	10	US-09-974-300-974
5	83	4.3	1131	10	US-09-974-300-5501
6	68	3.5	271	10	US-09-878-574-6807
7	60.6	3.1	1707	15	US-10-156-761-4723
8	60.6	3.1	9025608	15	US-10-156-761-1
9	59.2	3.1	1496	8	US-08-781-986A-274
10	56.2	2.9	380	9	US-09-770-791-195
11	54.4	2.8	1389	10	US-09-738-626-470
12	54.4	2.8	3309400	10	US-09-738-626-1
13	51	2.7	3303	15	US-10-156-761-5384
14	51	2.7	9025608	15	US-10-156-761-1
15	50.6	2.6	1965	13	US-10-190-435-50

16	50.6	2.6	1965	13	US-10-190-305A-55	Sequence 55, Appl
17	50.6	2.6	1971	13	US-10-190-435-49	Sequence 49, Appl
18	50.6	2.6	1971	13	US-10-190-305A-54	Sequence 54, Appl
19	50.6	2.6	2445	13	US-10-190-435-43	Sequence 43, Appl
20	50.6	2.6	2445	13	US-10-190-305A-37	Sequence 37, Appl
21	50.6	2.6	2457	13	US-10-190-435-44	Sequence 44, Appl
22	50.6	2.6	2457	13	US-10-190-435-45	Sequence 45, Appl
23	50.6	2.6	2457	13	US-10-190-305A-38	Sequence 38, Appl
24	50.6	2.6	2457	13	US-10-190-305A-39	Sequence 39, Appl
25	50.6	2.6	3162	13	US-10-190-435-18	Sequence 18, Appl
26	50.6	2.6	3231	13	US-10-190-435-60	Sequence 60, Appl
27	50.6	2.6	3231	13	US-10-190-305A-84	Sequence 84, Appl
28	50.6	2.6	3234	13	US-10-190-435-51	Sequence 51, Appl
29	50.6	2.6	3234	13	US-10-190-305A-56	Sequence 56, Appl
30	50.6	2.6	3462	13	US-10-190-435-16	Sequence 16, Appl
31	50.6	2.6	3531	13	US-10-190-435-13	Sequence 13, Appl
32	50.6	2.6	3537	13	US-10-190-435-14	Sequence 14, Appl
33	50.6	2.6	3537	13	US-10-190-435-15	Sequence 15, Appl
34	50.6	2.6	3597	13	US-10-190-435-46	Sequence 46, Appl
35	50.6	2.6	3597	13	US-10-190-305A-40	Sequence 40, Appl
36	50.6	2.6	3607	13	US-10-190-435-48	Sequence 48, Appl
37	50.6	2.6	3607	13	US-10-190-305A-42	Sequence 42, Appl
38	50.6	2.6	3624	13	US-10-190-435-47	Sequence 47, Appl
39	50.6	2.6	3624	13	US-10-190-305A-41	Sequence 41, Appl
40	50.6	2.6	3930	13	US-10-190-435-9	Sequence 9, Appl
41	50.6	2.6	3930	13	US-10-190-435-10	Sequence 10, Appl
42	50.6	2.6	3930	13	US-10-190-435-11	Sequence 11, Appl
43	50.6	2.6	4419	13	US-10-190-435-19	Sequence 19, Appl
44	50.6	2.6	4419	13	US-10-190-305A-14	Sequence 14, Appl
45	50.6	2.6	4713	13	US-10-190-435-59	Sequence 59, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-938-842A-1201  
; Sequence 1201, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 1201  
; LENGTH: 1632  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-1201

Query Match	35.3%	Score	679.6;	DB 10;	Length	1632;
Best Local Similarity	68.6%	Pred. No.	8.2e-177;			
Matches	985;	Conservative	0;	Mismatches	439;	Indels 12; Gaps 3;
Qy	235	CCGTGGCAAGGCTGCCAGCTGCCGCCCTCGAGCAGTTCAAGATAT---	CCGCCGACC	291		
Db	206	CTGGCTTAATGTCAGTAGCATCTCTGCTCTTGAACAACTCAAGAACTCTCAGTGATC	265			
Qy	292	GGTACATGAAGGAAGGAGTACCATTAGTGTGATAGGCTTCAGTGTTACACACACACAG	351			
Db	266	GATATCAAGGAAGGAAGGAGTATTGTTGATTGGACTTAGTATTACACACAGCTCTG	325			

Qy	352	TGAGATGCGTGA	AAAAACTTGTCTGTTGCGAGGAAC	TGTGGCCCGCGCTATTTC	AAAGAAC	411
Db	326	TTGAGATGCGTG	AGAGCGCTTGTCTATACCTGAAGCTGAATG	GCCAAGAGCTATTGCTGAAT	385	
Qy	412	TCACTAGCTGCA	ACCATTATTGAGAGAGCTGCTGTTCTTACTACTGCTGTAATAG	ATAGTGGAAA	471	
Db	386	TGTGTGTTTGA	ATCATATTGAGAGAGCTGCTGTGCTTTAGTACTTGTAA	CCGTATGGAGA	445	
Qy	472	TTTTATGTGTGG	CGCTATCATTGGAAACCGTGGTATCAGAGAAGTAGTGG	ACTGGATGTCGA	531	
Db	446	TTTTATGTTT	TAGCTTTATCTCAGCATCGTGGAGTTAAAGAAGTTACTGAAT	TGGATGTCAA	505	
Qy	532	AGAAAAAGTGTA	TATCCCGCTTCCGAGCTTAGGGAGCACCTGTTCACTCTTGG	AAGCAAGT	591	
Db	506	AGACGAGTGGG	ATCCCAAGTTTCAGAGATTTGTCAACACCGCTTTCTGTTGT	GTATCAACAACG	565	
Qy	592	ATGCCACACGC	ATCTGCTTTTGTAGGTGTCACTGGCCCTTGTACTCTTTTGGT	TCTCGGTGAAG	651	
Db	566	ATGCTACACAG	CATATATTTTGAAGTCTCAGCTGGTCTTGTACTCTCTCTGT	CCTAGGAGAG	625	
Qy	652	GACAAATCCTT	GTGCTCAGGTTAAACAAGTTGTGAGAGTGGACAGAA	CAGTGGAGGCTTGG	711	
Db	626	GTCAATCCTT	GCACAGGTGAACAAGTTGTGAAAGTTGTCAAGGAGTGAAT	TGGCTTTG	685	
Qy	712	GAAGAACA	TCGATAGATGTTCAAGATGCAATCACTGCTGGAAACGCTGT	TCGCAAGC	771	
Db	686	GGAGGAATAT	CAGCGGCTGTTTAAACACACGCGATAACTGTTGGT	TAAGCGTGTAGAACAG	745	
Qy	772	AGACCAACAT	ATCATCTGTGCTGTTTCTGTCACTCAGCGGGGTTTGAAT	CGCCCTGA	831	
Db	746	AGACAAATAT	TGCTTCTGTGCGGTTTCGGTTAGCTCAGCTGCTGTGTA	ACTTGTCTCTGA	805	
Qy	832	TGAAGCTTCC	GAAGTCTGAAAGCACTGTCACTAGGATGCTTCTGATTTGGT	GTGCTGTA	891	
Db	806	TGAAGCTTCC	GAGTCTTCAAA	CGTATCTGCTAGGATGTGTAAATTTGGCGCTGGA	865	
Qy	892	TGGGAAGCTAG	TGATCAAA	CATCTGGTTGCCAAGGATGCAGAAAGGTTGTTGTG	GTGA	951
Db	866	TGGGGAAGCT	TGTGATTAAGCATTTGATGGCTAAGGGTTGCACAAAAGT	GGTGGTAGTCA	925	
Qy	952	ACCCTCCGTG	GAAGGGTGGATGCTATTCTGTAGGAGATGA	AGATATAGAGATCTGT	1011	
Db	926	ACAGAAGTGA	AGAGGGTTTCACTATCCGTGAGGAGATGCCTGGT	TATTGAGATTAAT	985	
Qy	1012	ACAGGCTCTCT	CAGATGATPATAAGCTGCTGCTGAAGCTGATGTCGTGTT	TCACCAAGCA	1071	
Db	986	ATCGACCTCT	TGATGAGATGCTAGCTTGTGCTTCAGAAACGGA	TGTTGTTTACCAAGCA	1045	
Qy	1072	CCGCATCTGAA	CTTCAATGTTCCGAAGAAGACAGCGAGGCACTCC	CCCTGTCTCTG	1131	
Db	1046	CAGCCTCTGAG	ACGCCATTTGTTTAAAGAGAGCATGTAGAGAA	TCTCCCTCAAGCTTCTC	1105	
Qy	1132	ATACTATGGG	AGTGTTCGCTGTCACATATCTGTCCCAGGAATGT	CAGCGCAT	1191	
Db	1106	CAGAGGTTG	GAGGACTGAGGCATTTTGTGACATCTCTGTTC	CGAGAAATGTTGGATCTT	1165	
Qy	1192	GTGTGTCTGA	AGTTGGCGCTGCACGAGTGTCAATGTGACAGCTT	GAAGAAGTGGTG	1251	
Db	1166	GTGTGTGTGA	GTAGAAACACACGGGTTTACATGTGGACGAT	TCAAAAGAGTTCGTTG	1225	
Qy	1252	AAGCCAAACAG	GAGGACCGGCTCAGGAAGCAATGGAGGGCGCAG	CAATCATCACCGAAG	1311	
Db	1226	CTGCCAAACA	AGAAGACAGGATGAGGAAGCAATGGAAGCT	CAAA	1285	
Qy	1312	AAC	TGAGACGGTTCAGGCGATGGAGGACCTCGCTGGAG	ACGGTTC	1371	
Db	1286	AATCCACAC	AGTTTGAAGCATGGAGGGATTCATTGGAGA	CGGTTCCAA	1345	
Qy	1372	TGAGGTCGTAC	GCGGACAGGATCAGGCGCTCGGAGCTCGA	AGAGTGCCTCGA	1431	
Db	1346	TAAAGAGCTT	ATGCAGAGAGAAATCAGAGTGCAGAGCTCGA	AGAGTGCATGTCCAAATGG	1405	

Qy	1432	GTGAGAGCGCCCTCACAAGAGATGAGAGAGCCATCGAGGAGCTTGAGCACCGGCATCG	1491
Db	1406	GAGATGA---TATCAACAAGAAACAACAGAGAGCGTGTGATGATTTAAGCCGAGGTATCG	1462
Qy	1492	TTAACAGAGCTCTCCATGGCCGCTGCAGCACCTGAGGTGCGACGCGCAGCAGCGCA	1551
Db	1463	TGAACAGATTCTTACATGGTCCATCGCATTTTGAGATGTGACGGAAGTGACGTAGAA	1522
Qy	1552	CCCTTGACGAGAGCGCTCGAGAACATCAACGCCCTCAACGGGATGTTTCAGCCTCGACATCG	1611
Db	1523	CGCTGAGCGAGACCCCTTGAGATATCATGCTTTGAACAGAAATGATCGGTC-----TTG	1576
Qy	1612	AGAAGCGGATCATCGAGCGAGAGATCAAGGCCAAGGTGAGAGAGACACAAAACCTGA	1667
Db	1577	AGAAAGACATTTTGGAGGAAAGCTTTAAGGCAATGCGGAAACAACAACAGAAGTAA	1632

RESULT 2  
 US-09-878-574-3208  
 ; Sequence 3208, Application US/09878574  
 ; Patent No. US20020110548A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Byrum, Joseph R.  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Thompson, Michael D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(15401)B  
 ; CURRENT APPLICATION NUMBER: US/09/878,574  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 09/333,535  
 ; PRIOR FILING DATE: 1999-06-14  
 ; NUMBER OF SEQ ID NOS: 15775  
 ; SEQ ID NO 3208  
 ; LENGTH: 384  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; OTHER INFORMATION: Clone ID: LIB3028-014-Q1-B1-D9  
 US-09-878-574-3208



```

: GENERAL INFORMATION:
: APPLICANT: Byrum, Joseph R.
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Thompson, Michael D.
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(15401)B
: CURRENT APPLICATION NUMBER: US/09/878,574
: CURRENT FILING DATE: 2001-12-21
: PRIOR APPLICATION NUMBER: 09/333,535
: PRIOR FILING DATE: 1999-06-14
: NUMBER OF SEQ ID NOS: 15775
: SEQ ID NO 5465
: LENGTH: 275
: TYPE: DNA
: ORGANISM: Glycine max
: OTHER INFORMATION: Clone ID: 701097007H1
: US-09-878-574-5465

```

Query Match	6.4%	Score 122.6;	DB 10;	Length 275;
Best Local Similarity	71.4%	Pred. NO. 2.2e-23;		
Matches 175;	Conservative	0;	Mismatches 69;	Indels 1; Gaps 1;

603	Qy	CATCTGTTT	GAGGTGCAGCTGGCC	TTGACTCTTTGGTTCTCGGTGAAGGACAATCCCTT	662
7	Db	CATCTATTT	GAAATGGCGTCAGGGCTTGACTC	TGTCTTTGGGGAAGGTCAAATTCCTT	66
663	Qy	GCTCAGGTTAA	CAAGTTGTGAGGAGTGGACAGAACAGTGGAGCGT	TTGGGAAAGAACATC	722
67	Db	GCTCAGTCA	GACGAGTTGTGAACTGGA	CAGGGAGTGCTGTTTGTATAGAAATC	126
723	Qy	GATAGATGTT	CAAGATGCAATCACTGCTGGAAAGCGTGTCCGACGAGACCAACATA	782	
127	Db	AGTGTTT	CTCAAGCAGCGCATATCGGTTGGGAAGCGGGTTAGAACCGAGACTAACATT	186	
783	Qy	TCATCTGGNGCT	GTGTTCTGTCAAGTTACAGCGCGGTTGAACTGGCCCTGATGAAGCTTCCG	842	
187	Db	TATCTGGATCA	GTTTCTGTAAAGCTCGCGTCTGTGGAGCT-GCACTGATGAAGCTACCG	245	
843	Qy	AAGTC	847		
246	Db	GATTC	250		

```

RESULT 4
US-09-974-300-974
; Sequence 974, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 974
; LENGTH: 1360
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-974

```

Query Match 4.5%; Score 86.6; DB 10; Length 1360;  
Best Local Similarity 44.7%; Pred. No. 5e-13;  
Matches 485; Conservative 0; Mismatches 579; Indels 21; Gaps 3;  
QY 315 ATAGCTGTGATAGGCCTCAGTGTGTACACACAGCACCAGTGGAGATGCGTGAANAACCTTGCT 374

Db	7	AT	CT	CT	CT	CG	CG	GA	TT	GG	AA	TT	CA	AA	T	C	A	G	C	C	C	T	TT	G	A	T	A	C	G	T	G	A	A	A	G	T	T	A	A	C		66										
Q1	375	GT	TC	G	A	G	A	A	C	T	G	T	G	C	C	C	G	T	G	T	A	T	T	C	A	A	G	A	A	C	T	C	A	T	G	A	C	C	T	G	A	A	C	434								
Db	67	TT	T	C	A	G	C	C	G	A	T	G	A	G	C	T	G	G	C	G	A	T	T	G	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	126										
Q1	435	G	A	G	C	T	C	T	G	T	T	C	T	T	A	G	T	A	C	T	G	T	A	T	A	G	A	A	T	T	T	A	T	G	T	G	T	G	G	C	G	C	A	T	G	494						
Db	127	G	A	G	A	A	C	A	T	C	A	T	C	G	T	G	C	A	T	G	A	C	C	G	A	C	A	G	A	G	C	T	G	T	A	G	A	C	T	G	T	G	A	C	G	186						
Q1	495	A	A	C	C	G	T	G	T	A	T	C	A	G	A	A	G	T	A	G	T	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	554							
Db	187	C	A	C	C	G	A	C	G	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	A	246							
Q1	555	G	A	G	C	T	T	A	G	G	A	G	A	C	C	T	G	T	C	A	T	T	C	G	A	A	G	A	G	A	T	A	T	C	C	A	C	A	C	C	A	T	C	T	G	T	T	G	A	614		
Db	247	G	A	C	A	T	T	T	C	G	C	T	T	A	C	T	T	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	A	A	306					
Q1	615	G	T	G	T	C	A	G	C	T	G	C	T	T	G	A	C	T	T	T	T	T	C	G	T	G	A	A	G	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	674					
Db	307	G	T	A	T	C	G	T	C	G	A	T	T	G	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	366					
Q1	675	C	A	A	G	T	T	G	T	G	A	G	A	T	G	G	A	C	A	G	T	G	G	A	G	C	T	T	C	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	734					
Db	367	A	C	A	A	G	T	T	T	A	A	G	T	T	C	A	G	A	A	G	A	A	G	A	A	T	T	G	T	A	C	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	426					
Q1	735	A	A	G	A	T	C	A	T	C	A	T	G	T	G	A	A	A	A	G	T	G	T	C	C	G	A	G	A	C	C	A	A	C	A	T	A	T	C	A	T	C	A	T	C	T	G	T	G	C	T	794
Db	427	A	A	G	C	A	G	C	G	T	C	A	C	G	T	C	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	486					
Q1	795	G	T	T	C	T	C	A	G	T	T	C	A	G	C	G	C	G	T	T	G	A	A	C	T	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	854					
Db	487	G	T	A	T	C	C	G	T	C	A	G	T																																							

```
Db      1066 CAGGC 1070

RESULT 5
US-09-974-300-5501
; Sequence 5501, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5501
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-5501

Query Match      4.3%; Score 83; DB 10; Length 1131;
Best Local Similarity 46.9%; Pred. No. 4.5e-12;
Matches 380; Conservative 0; Mismatches 410; Indels 21; Gaps 3;

QY      555 GAGCTTAGGAGACCTGTCATCTTGGGAGCAGTGATGCCACAGCCATCTGTTGAG 614
Db      47 GAGCTAACGCCGATTTGGTTATTCCTGGAATGACCATGCCATTAACACCTTTCCGC 106
QY      615 GTGTACAGTGGCCTTGACTCTTTGGTTCTCGGTGAAGGACAAATCTTGTCAGGTTAA 674
Db      107 GTTCTTGTGGTTAGACTTATGTTGATCGGGAACGCAATTTCTTGACAGTGAAG 166
QY      675 CAAAGTTGAGGAGTGGACAGAAACAGTGAGGCTTGGGAAGAACATCATATATCTGTGTC 734
Db      167 CAGGCATTTTGTGTTGCCCAAGAGAAAGTACAGGCACTGTGTTTAAACAGCTATTT 226
QY      735 AAGGATGCAATCACTGCTGGAAGCTGTCGGCAGGACCAACATATATCTGTGTC 794
Db      227 AAACAAGCGGTACGCTCGCAACAGTGTGCACTCGTCAACGGAATTTCTTCAACAGCC 286
QY      795 GTTCTGTGTCAGTTTCAGCGCGGTTGAACCTGGCCCTGATGAAGCTTCCGAAGTCTGAAGCA 854
Db      287 GTTCTGTGTCATGCTATGACGCGGTAGAACTGGGAAGAAATTTTCGCGACGTTTTCAGGA 346
QY      855 CTGTACAGTAGGATGCTTCTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 914
Db      347 AAACATG-----TGCTTATTTTAGCGCTGGAAAAATGAGCGAACTGACGCGCAACAT 400
QY      915 CTGGTTGCCAAAGGATGCAAGAGTTGTTGTGTCACCGCTCCGTCGGAAGGGTGGAT 974
Db      401 CTTTATTCGATGCGGCTGCTTCCATAGTCAATGACCGCAACGGAAGAAAGAAAGCGGTG 460
QY      975 GCTATTCTGTGAGGAGTGAAGATATAGAGATCGTGTACAGGCTCTCTCAGACATGTAT 1034
Db      461 GAACTGGCAAGCCAGTTTCTGGAACGGCT-----CGGTCTTTTGTGTAATTGAT 511
QY      1035 CAACTGCTGCTGAAGCTATGTCGTTGTTACACGACCGCATCTGAAACTTCATGTTGTC 1094
Db      512 GATGCTTTTAAAGAACGATGTTGATAGTCTCACCGGGGCCCGCGACTTTGTTGTC 571
QY      1095 GCAAAAGAACACGACGAGGCACTCCCGCTGCTCTGATATGATGAGTGTGTCGCTG 1154
Db      572 AAAAGGAATATGCGCAGCTGCC-----TTAAAAACGGAAGGAGCGCGCTGTTT 625
QY      1155 TTTGTGCAATATCTGTCCCGAGGAATGTGAGGCAATGTGTCGAAAGTGTGCGCTGCA 1214
```

```
Db      626 GCGATTGATATCGCAGTGCCACGGGATATTGACCCGGAATTTGCAACAATTAGCGAGCTT 685
QY      1215 CGAGTTACAAATGTCGACGACTTTGAAAGAGGTGGTGGAGCCAAAGAGGACCGGCTC 1274
Db      686 TACTTTACGACATTTGATGACTTTACAGAACATTTGTTGAAGCCCAATAAAAAAGAGCGCAAA 745
QY      1275 AGAAAGCAATGGAGCGCGACAGACATCATCAGGAGAACTGAGACGGTTTCGAGGCAATGG 1334
Db      746 AAAGAGGCTGAAAAAATCGGTGTTATGATTGAGGAGAGATCGATGACTTTTAAAGCATGG 805
QY      1335 AGGGAATCGCTCGAGACCGTTCCGACCATCA 1365
Db      806 CTAACACACTAGGCGGTGTGCTCTTATTA 836

RESULT 6
US-09-878-574-6807
; Sequence 6807, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 6807
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701099031H1
US-09-878-574-6807

Query Match      3.5%; Score 68; DB 10; Length 271;
Best Local Similarity 62.2%; Pred. No. 2.7e-08;
Matches 107; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY      1494 ACACAGCTCTCCATGGCCCGCTGACGACCTTGAGTGGCGGACGACAGCCGACC 1553
Db      10 AATAAGCTACTTCTGTTGTTCCCATGCGACACCTTAAGGTGTGATGGGAACAATGATAGT 69
QY      1554 CTTGACGAGACGCTCGAGAACATGACGCGCTCAACCGGATGTTCAAGCTTCGACATGGAG 1613
Db      70 CTGAGTGAAGTACTTTGAGNATATGCGCCCTTACAGAAATGTATGATCTTGACAGAA 129
QY      1614 AAGGCGATCATCGACAGAAAGATCAAGGCCAAGGTGGAGAGACACAAACT 1665
Db      130 ACTTCTTGATCGAAGAAAGATCAGAGTCAAGATGGAACGGGTTTCAAGAT 181

RESULT 7
US-10-156-761-4723
; Sequence 4723, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
```

;; PRIOR APPLICATION NUMBER: JP 2001-272697  
;; PRIOR FILING DATE: 2001-08-02  
;; NUMBER OF SEQ ID NOS: 15109  
;; SEQ ID NO 4723  
;; LENGTH: 1707  
;; TYPE: DNA  
;; ORGANISM: Streptomyces avermitilis  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)..(1707)  
US-10-156-761-4723

Query Match  
Best Local Similarity 3.1%; Score 60.6; DB 15; Length 1707;  
Matches 168; Conservative 0; Mismatches 179; Indels 0; Gaps 0;  
QY 435 GAGGCTGCTTCTTAGTACCTGTATAGAAATTTATGTGTGGCGCTATCATGG 494  
DB 127 GAGGCGCGGCTCTTGCCACCTGCAACCGCATCGAGCTGTACGCCAGCTGACAAATTC 186  
QY 495 AACCGTGTATCAGAGAACTAGTGGACTGTGATGTCGAAGAAAGTGGTATTCCTCGCTTCC 554  
DB 187 CACGCGGCGTGGCCGAGCTGTCCAGCTGTCTGCCACACACAGCGGGGTCTGGGCTCGAG 246  
QY 555 GAGCTTAGGAGACCTGTGTCATCTTTGCGAAGCAGTGTGCCACACGCCCATCTGTTTGA 614  
DB 247 GAGCTCACTCCCTATCTCTACGTGCACTACGAGGACCGGGCGCTCCACCACTCTTCTCG 306  
QY 615 GTGTACGTGGCTTGGACTCTTTGTTCTCGGTGAAGCAAAATCTTCTCAGGTTAAA 674  
DB 307 GTGGCTGCGGCTCGACATGATGTTGTTGCGGCGAGGCGCAGATCTCTGCCAGATCAAG 366  
QY 675 CAAGTTGTGAGGAGTGGACAGAACAGTGGAGGCTTGGGAAAGAAAGATCGATAGGATGTC 734  
DB 367 GAGCGCTGGCCACCGCGAGAACTGACACCGGGGACGGCTCTCAACGACCTGTTTC 426  
QY 735 AAGGATGCAATCACTGTGCGAAAGCTGTCCGACGCGAGACCAACAT 781  
DB 427 CAGCAGGCACTGCGGACCGGCAAGCGGGCGCACTCCGAGACCGGCAT 473

## RESULT 8

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown

Query Match  
Best Local Similarity 3.1%; Score 60.6; DB 15; Length 9025608;

Best Local Similarity 48.4%; Pred. No. 0.0014;  
Matches 168; Conservative 0; Mismatches 179; Indels 0; Gaps 0;  
QY 435 GAGGCTGCTTCTTAGTACCTGTATAGAAATTTATGTGTGGCGCTATCATGG 494  
DB 5777935 GAGGCGCGGCTCTTGCCACCTGCAACCGCATCGAGCTGTACGCCAGCTGACAAATTC 5777994  
QY 495 AACCGTGTATCAGAGAACTAGTGGACTGTGATGTCGAAGAAAGTGGTATTCCTCGCTTCC 554  
DB 5777995 CACGCGGCGTGGCCGAGCTGTCCAGCTGTCTGCCACACACAGCGGGGTCTGGGCTCGAG 5778054  
QY 555 GAGCTTAGGAGACCTGTGTCATCTTTGCGAAGCAGTGTGCCACACGCCCATCTGTTTGA 614  
DB 5778055 GAGCTCACTCCCTATCTCTACGTGCACTACGAGGACCGGGCGCTCCACCACTCTTCTCG 5778114  
QY 615 GTGTACGTGGCTTGGACTCTTTGTTCTCGGTGAAGCAAAATCTTCTCAGGTTAAA 674  
DB 5778115 GTGGCTGCGGCTCGACATGATGTTGCGGCGAGGCGCAGATCTCTGCCAGATCAAG 5778174  
QY 675 CAAGTTGTGAGGAGTGGACAGAACAGTGGAGGCTTGGGAAAGAAAGATCGATAGGATGTTTC 734  
DB 5778175 GACGCGTGGCCACCGCGAGGAACTGACACCGCGGGGACGGCTCTCAACGACCGGCAT 5778234  
QY 735 AAGGATGCAATCACTGTGCGAAAGCTGTCCGACGCGAGACCAACAT 781  
DB 5778235 CAGCAGGCACTGCGGACCGGCAAGCGGGCGCACTCCGAGACCGGCAT 5778281

## RESULT 9

US-08-781-986A-274  
; Sequence 274, Application US/08781986A  
; Publication No. US20030054436A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781.986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 274:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1496 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-781-986A-274

Query Match  
Best Local Similarity 3.1%; Score 59.2; DB 8; Length 1496;  
Matches 185; Conservative 1; Mismatches 184; Indels 6; Gaps 1;



QY 703 GAGGCTTGGAAAGAACATCGATAGATGTTCAAGATGCAATCACTCTCGAAGCGTG 762  
 Db 395 GTTCTGTCGTCCTCTTTCATTCATTCACCAACCGCGTGCATACCGCAAGCGG 454  
 QY 763 TCCGAGCGAGACCAACATATCATCTGTCGTCGTTTCTGTCAGTTTCAGCGCGGTTGAAC 822  
 Db 455 TGCATTGCGAGACTCTGATTGATGATGCTGGTGCATCGATGCTGTCCTTTCGCTGTGGATC 514  
 QY 823 TGGCCCTGATGAAGCT 838  
 Db 515 GCGCGTTGGTCAGAT 530

RESULT 12

US-09-738-626-1  
 ; Sequence 1, Application US/09738626  
 ; Publication No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOGUCHI, HIROSHI  
 ; APPLICANT: ANDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIKO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAOKO  
 ; APPLICANT: SENOH, AKIHIRO  
 ; APPLICANT: IKEDA, MASATO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; CURRENT FILING DATE: 2000-12-18  
 ; PRIOR APPLICATION NUMBER: JP 99/377484  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: JP 00/159162  
 ; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: JP 00/280988  
 ; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 7059  
 ; SOFTWARE: Patent in ver. 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 3309400  
 ; TYPE: DNA  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-738-626-1

Query Match 2.8%; Score 54.4; DB 10; Length 3309400;  
 Best Local Similarity 44.4%; Pred. No. 0.041;  
 Matches 220; Conservative 0; Mismatches 276; Indels 0; Gaps 0;  
 QY 343 CAGCACCAGTGGAGATGGTGAAGAACTTGTCTGTCAGAGGAATGTGGCCCGTGTGTA 402  
 Db 438550 CTGCGCTGTGTCGCTTCTTGAACGCTCTGAGCATGGATGATTCAGTACGTGGTGAACAA 438609  
 QY 403 TTCAAGAACTCACTAGCTGAACCATATTTGAAGAGGCTGCTGTTCTAGTACCTGTAATA 462  
 Db 438610 CTCAGGACTCTCTGGTAGGCGCTTTTAAAGGAGGCGCTCATTTCTTACGTGTAAAC 438669  
 QY 463 GAATGGAATTTATGTGTGGCGGTATCATGAAACCGTGTGATCAGAGAACTAGTGGACT 522  
 Db 438670 GCCTGGAGGTCTACACCGTCACTAGCAGTTTCCATACCTGTTGTTAATGATGTGTGGAGG 438729  
 QY 523 GGATGTCGAAGAAAGTGTATTCGCCGTTCCGAGCTTAGGAGCACTGTTCAATCTTTC 582  
 Db 438730 TTCTCCATGAGGCAAGTGGCGTAGATATTGAACCTTTTGGCGGATATCTTTATGTCGTT 438789  
 QY 583 GAAGCAGTGTGCCACACGCCATCTGTTTGAAGTGTGTCAGCTGGCTTGACTCTTTGGTTC 642  
 Db 438790 AGCCGATGCTGCTGCTGAACACATGTTGGTGGTACTTCGGGTGGATTCATCTGTTGT 438849  
 QY 643 TCGGTGAAGACAAATCTTCTGCTCAGGTTAAACAAAGTTGTGAGGAGTGGACAGAACAGTG 702

Db 438850 TGGGTGACGACGATCATTCGTGAGTGGCGCACTGCGTACCAAGCAGCTAATGATATG 438909  
 QY 703 GAGGCTTGGAAAGAACATCGATAGGATGTTCAAGATGCAATCACTCTGTCGAAAGCGTG 762  
 Db 438910 GTTCTGTCGTCCTCTTTCATTCATTCACCAACCGCGTGCATACCGCAAGCGG 438969  
 QY 763 TCCGAGCGAGACCAACATATCATCTGTCGTCGTTTCTGTCAGTTTCAGCGCGGTTGAAC 822  
 Db 438970 TGCATTGCGAGACTCTGATTGATGATGCTGGTGCATCGATGCTGTCCTTTCGCTGTGGATC 439029  
 QY 823 TGGCCCTGATGAAGCT 838  
 Db 439030 GCGCGTTGGTCAGAT 439045

RESULT 13

US-10-156-761-5384  
 ; Sequence 5384, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: HORIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 5384  
 ; LENGTH: 3303  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(3303)  
 US-10-156-761-5384

Query Match 2.7%; Score 51; DB 15; Length 3303;  
 Best Local Similarity 50.6%; Pred. No. 0.0058;  
 Matches 123; Conservative 0; Mismatches 120; Indels 0; Gaps 0;  
 QY 72 ATGCGGACGACGACGTGAGGACCAACCGCGCGCGAGCGCCACCCAGCGCGGCGGCGGCGGCGG 131  
 Db 433 ATGCGGCTGACGACGACCGCGCGCGAGCGCGGACACGCGGCGGCGGCGGCGGCGGCGG 492  
 QY 132 CGGCGGTCGTCGTCGCGGCGCTCTCCAGAGGGTGGCGCGCGCGGCGGCGGCGGCGGCGG 191  
 Db 493 ACGGCGACGACTCTCCGCGCGCTCAAGAGCACCCCGCGCGGAGAACCCGCTACACCCCGTCC 552  
 QY 192 GTGTGTCGTCGACACGCGCGCGCTGAGCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGGTC 251  
 Db 553 TTGAGACGCGCGCGCTGCGACTTCTGTAAGGACCAACCCGAGGCGGCGGCGGCGGCGGCG 612  
 QY 252 AGCGTCGCCCGCTCTGAGCAGTTCAAGATATCCGCGCGCGCGCGCGCGCGCGCGCGGAGG 311  
 Db 613 ACCATCGCGCTCTCTGACTCGGCGGTGGACCTCGGCGCACCGCGCGGCTCCAGAACGACG 672  
 QY 312 ACC 314  
 Db 673 ACC 675

RESULT 14

US-10-156-761-1/c  
 ; Sequence 1, Application US/10156761

Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1  
LENGTH: 9025608  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4187715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 2.7%; Score 51; DB 15; Length 9025608;  
Best Local Similarity 50.6%; Pred. No. 0.64;  
Matches 123; Conservative 0; Mismatches 120; Indels 0; Gaps 0;  
QY 72 ATGGGACACACGAGCTCAGCGACACCGCCGCGGAGAGCGCCACACCGCCCAAGCGG 131  
Db 6549578 ATCGGCTCGAGACCGCGCGCGGAGCGCCGACCGCAAGGCGCGTCTCGAAGGCC 6549519  
QY 132 CGGGGGTCTGCTGGCCCTCTGCCAGAGGTGGCGCGGCGGAGCGCGCGTCCGGG 191  
Db 6549518 ACGGGCACGTACTCGCGCCGCTCAAGAGCACCCCGCGGAGAACCCGTACAAACCGTCC 6549459  
QY 192 GTGGTGGGTGGAGCGCGCGCGGCGGAGGCGCGGAGCGCGCGTGGCAAGGCTGCC 251  
Db 6549458 TTCGAGACGGGCGCGCTGACTTCGTGAAGGACCAACCGAAGCGGACGCGCGGCGCATC 6549399  
QY 252 AGCGTCGCGCGCTCGAGCAGTTCAAGATATCCGCGACCGGTACATGAAGGAGGAGT 311  
Db 6549398 ACCATCGGCTCTCTGACTCGGCGTGGACCTCGGCGCCCGCGCTCCAGAAGACGACC 6549339  
QY 312 ACC 314  
Db 6549338 ACC 6549336

RESULT 15  
US-10-190-435-50  
Sequence 50, Application US/10190435  
Publication No. US20030143248A1  
GENERAL INFORMATION:  
APPLICANT: ZUR MEGEDE, Jan  
APPLICANT: BARNETT, Susan W.  
APPLICANT: LIAN, Ying  
APPLICANT: ENGELBRECHT, Susan  
APPLICANT: VAN RENSBURG, Estrelita J.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
FILE REFERENCE: PP18133.003 / 2302-18133  
CURRENT APPLICATION NUMBER: US/10/190,435  
CURRENT FILING DATE: 2002-12-30  
NUMBER OF SEQ ID NOS: 319  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 50  
LENGTH: 1965  
TYPE: DNA

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: protInaRT.YMM.opt\_C  
US-10-190-435-50

Query Match 2.6%; Score 50.6; DB 13; Length 1965;  
Best Local Similarity 47.4%; Pred. No. 0.0055;  
Matches 152; Conservative 0; Mismatches 169; Indels 0; Gaps 0;  
QY 1334 GAGGACTCGCTGGAGACCGTTCCGACCATCAAGAAGCTGAGGTCTGTACGCGGACAGGAT 1393  
Db 1560 GTGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGACGGCGC 1619  
QY 1394 CAGGCGCTCGGAGCTCGAGAAGTGCCTCGAGAAGTAGTGTAGGAGCGCCCTCACCAAGAA 1453  
Db 1620 CGCCAAACCGGAGAGCAAGATCGGCAAGGCCGCTACGTGACCGACCGGGCGGCGAGAA 1679  
QY 1454 GATGAGGAGAGCCATCGGAGGAGCTGAGCACCGGCAATCGTTAACAGCTCTCCATGGCCC 1513  
Db 1680 GATCGTAGCCTGACCGGAGACCCACACAGAGACCGAGCTGCGAGGCGCATCCAGCTGGC 1739  
QY 1514 GCTGAGCAGCTGAGGTGCGAGCGGAGCGAGCAGCGCCACCTTTGACGAGACGCTCGAGAA 1573  
Db 1740 CCTGAGGAGACAGCGGAGGTTGAACATCGTGACCGACAGCCAGTACGCCCTGGCAT 1799  
QY 1574 CATGACCGCCCTCAACCGGATGTTTTCAGGCTCGACATGAGAGAGGCGGATCATCGAGCAGAA 1633  
Db 1800 CATCCAGGCCCGCGCCGCAAGAGCGAGGAGCGGCTGGTGAACCAAGATCATCGAGCAGCT 1859  
QY 1634 GATCAAGGCCAAGGTGGAGAA 1654  
Db 1860 GATCAAGAGGAGAGGTGTA 1880

Search completed: December 15, 2003, 18:32:45  
Job time : 1550 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 12:46:10 ; Search time 111 Seconds  
(without alignments)  
7650.641 Million cell updates/sec

Title: US-10-018-902-3

Perfect score: 1924

Sequence: 1 ccacgcgtccgcatcaataa.....aaaaaaaaaaaaaaaaaag 1924

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*

2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*

3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*

4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*

5: /cgn2\_6/prodata/2/ina/PCFUS\_COMB.seq:\*

6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87.4	4.5	1308	4	US-09-252-991A-13991
2	87.4	4.5	1338	4	US-09-252-991A-14144
3	82.6	4.3	1026	4	US-09-252-991A-14075
4	64.6	3.4	1347	4	US-09-134-001C-218
5	64.2	3.3	1664976	4	US-08-916-421B-1
6	57.4	3.0	1290	4	US-09-328-352-517
7	56.4	2.9	7218	1	US-08-232-463-14
8	53	2.8	4403765	3	US-09-103-840A-2
9	53	2.8	4411529	3	US-09-103-840A-1
10	50	2.6	8147	4	US-09-514-247A-9
11	49.6	2.6	42988	4	US-08-311-731A-128
12	49.4	2.6	1203	3	US-09-086-010-1
13	48.8	2.5	1149	2	US-08-466-103A-3
14	48.4	2.5	77536	4	US-09-410-551B-1
15	47.6	2.5	4403765	3	US-09-103-840A-2
16	47.6	2.5	4411529	3	US-09-103-840A-1
17	45	2.3	768	4	US-09-252-991A-12559
18	45	2.3	1377	3	US-09-377-557-17
19	45	2.3	1425	4	US-09-252-991A-12926
20	45	2.3	2002	4	US-09-016-434-1172
21	44.6	2.3	984	4	US-09-252-991A-14047
22	44.6	2.3	1488	4	US-09-252-991A-8426
23	44.6	2.3	2049	4	US-09-252-991A-8348
24	44.6	2.3	2118	4	US-09-252-991A-14221
25	44.6	2.3	2121	4	US-09-252-991A-14004
26	44.6	2.3	2205	4	US-09-252-991A-14132
27	44.4	2.3	5118	4	US-08-669-785-3

C 28	44.2	2.3	1935	4	US-09-620-312D-236	Sequence 236, Appl
C 29	44.2	2.3	2825	4	US-09-196-390-5	Sequence 5, Appli
C 30	43.8	2.3	1134	4	US-09-622-439-21	Sequence 21, Appli
C 31	43.8	2.3	1512	3	US-08-476-509B-1	Sequence 1, Appli
C 32	43.8	2.3	1626	3	US-08-348-518C-1	Sequence 1, Appli
C 33	43.6	2.3	744	4	US-09-368-819A-1	Sequence 3, Appli
C 34	43.6	2.3	795	4	US-09-368-819A-3	Sequence 3, Appli
C 35	43.6	2.3	1173	3	US-08-993-380-3	Sequence 3, Appli
C 36	43.4	2.3	1621	1	US-08-722-001-13	Sequence 13, Appli
C 37	43.4	2.3	4079	4	US-09-016-434-1237	Sequence 1237, Ap
C 38	43.2	2.2	2175	4	US-09-252-991A-15144	Sequence 15144, A
C 39	43.2	2.2	2588	2	US-08-796-414B-6	Sequence 6, Appli
C 40	43.2	2.2	2721	4	US-09-252-991A-14887	Sequence 14887, A
C 41	43.2	2.2	2781	4	US-09-252-991A-14482	Sequence 14482, A
C 42	43	2.2	494	4	US-09-056-556-176	Sequence 176, App
C 43	43	2.2	494	4	US-09-072-596-171	Sequence 171, App
C 44	43	2.2	1647	1	US-08-083-948-13	Sequence 13, Appl
C 45	43	2.2	1647	1	US-08-393-785-13	Sequence 13, Appl

## ALIGNMENTS

### RESULT 1

US-09-252-991A-13991  
; Sequence 13991, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13991  
; LENGTH: 1308  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13991

Query Match	4.5%	Score 87.4;	DB 4;	Length 1308;
Best Local Similarity	46.5%;	Pred. No. 2e-10;		
Matches 585;	Conservative	0;	Mismatches 636;	Indels 36; Gaps 8;
QY	313	CCATAGCTGTGTAGCCCTCAGTGTACACAGCAGCAGTGGAGATGCGTGAACAACTTG	372	
DB	44	CCTTCATTGCCCTCGGCATCAATCAAGACCGCCTCGGTGGCTGTCGCGAGCGGTGG	103	
QY	373	CTGTGTGAGAGAACTGTGGCCCGCTGCTTCAAGAACTCACTAGCCCTGAACCATATTG	432	
DB	104	CTTTCACCTCCGAAAGATGCTGAGGCGCTGACAGCTGTGCCCTGACACACAGCC	163	
QY	433	AAGAGCTGCTGTTCTTAGTACCTGTAATAGAAATTTATGTGTGGCGCTATCAT	492	
DB	164	GCGAGGCGGCCATCTGTGACCTGCAACCGAGCGAACTCTACCTGGAGAT-CGACCAT	222	
QY	493	GGAACCGTGGTATCAGAGAGTAGTGGACTGATGCGAAGAAAGTGGTATTCGCGCTT	552	
DB	223	CCGACCGCGGAC-----GACGTGCTCGCTGCTGCGGACTACCATCGCTGACTCTCG	277	
QY	553	CCGAGCTTAGGAGCAGCACCTGTTTCATCTTCCGAAGCAGTATGCCACGCCCATCTCTT	612	
DB	278	ACGAGTTGGCGCCCTGTGCTTATGTGCACAGGAGAGACGCGGTGCCACATATGTC	337	
QY	613	AGGTGTAGCTGGCCTTGACTCTTTGGTTCTCGGTGAAGGACAAATCTTGTTCAGGTTA	672	
DB	338	GCGTGGCCTCGGCCCTCGACTCGATGGTCTCTCGGAGCGCGCAGATCTCGCCAGATCA	397	



```
QY 673 AACAGTTGTGAGGAGTGGACAGACAGTGGAGGCTTGGAAAGAACATCATAGATGT 732
Db 398 AGTCGCGCTACGCGGTGCGCGGAGGCGGACCGTCTGCTGCTGCTGCGCGCGCTGT 457
QY 733 TCAAGGATCAATCACTGCTGGAAGCGTTCGCGAGCGAGACCAACATATCATCTGGTG 792
Db 458 TCCAGGCCACCTTCAGTACGCGGAGAGACCGTCCGACCGATACCGCATCGCGAGAAC 517
QY 793 CTGTTTCTGTCACTCAGCGCGGTGAACTGCGCCCTGATGAAGCTTCCGAAGTCTGAAG 852
Db 518 CGGTGTCGCTGGGCTGCGCGCGGTGAGCTGCGCCCAAGCAG-ATCTTCAGCGACCTGCAC 576
QY 853 CACTGTCACTAGGATGCTTCTGATTGGTCTGCTAAATGGAAAGCTAGTATCAAAAC 912
Db 577 CGAGCCAGGC-----GCTGCTGATCGCGCCGCGAGACCATCACCTGCTCGCGGCC 631
QY 913 ATCTGTTGCCAAGAGATGCAAGAGTTGTTGTGTGAACCGCTCCGTTGGAAGGTTGG 972
Db 632 ACCTGTTTCGAGCAAGCGTGAAGCGATCGTGTGGCCCAACCGCACCTGGAGCGGCCA 691
QY 973 ATGCTATTCTGAGGAGATGAAGATATAGAGATCGTGTACAGGCTCTCTCAGACATGT 1032
Db 692 GCTGCTCGCGAGCAGTTGCGCGCCCATGCGGTGCTG-----CTTTCGAGATAC 742
QY 1033 ATCAAGCTGCTGTAAGCTGATGTGCTGTTCACAGCACCGCATCTGAAGCTTCAATTGT 1092
Db 743 CGAGGAGTTGGCCACAGCGATATGTCATAGCTCCATGTCAGCGCATGTCGCCATCC 802
QY 1093 TCGCAAAAGAACACGAGAGGCACTCCCCCTGCTCTGTATATATATGGAGGTGTTGCGC 1152
Db 803 TCGCAAGGCGCGCTGGAG-CGCGGCTGGAAGCGCAAGCAAGCGATGTTCA-- 859
QY 1153 TGTTCGTGACATATCTGTCCCGAGGAATGTCAGCATGTGTCTGAAGTTGGCGCTG 1212
Db 860 ---TGTTCGACATCGCGCTGCGCGGACATCGAGCGGAACTGCGCGAACTGGACGAG 916
QY 1213 CACGAGTGTACAAATGCGACGACTTGAAGAGGTGTGGAAGCAACAAAGGAGGACCGGC 1272
Db 917 TCTACCTCTATACGCTCGAGCACTCCACAGATAGTGGCGGAGATCTCAAGAGCGCC 976
QY 1273 TCAGGAAGCAATGGAGGCGCAGACATATCAACGAAGACTGAGACGCTTCGAGGCAT 1332
Db 977 AGGCGCTGCGCAAGGCGCGGAGAACTGCTGCGCAGCGCGCTGCGCGATTCATGACG 1036
QY 1333 GGAGGACTCGCTGGAGACCGTTCCGACCATCAAGAGCTGAGTCTGACGCGGACAGGA 1392
Db 1037 GCTTCGCGMACTGGCGCGGTGGAAGTGTGTCGCGCGCTATCGCGAGCGCGCGGCC 1096
QY 1393 TCAGGCGCTCGGAGCTCGAAGTGTCTCGAAGAGTGTGAGGACGCGCCCTCACCAAGA 1452
Db 1097 TCGCGATGAGGAGCTGGGCAAGCGCCACGCGCACTGGCCACGCTGCGACCCCGCG 1156
QY 1453 AGATGAGGAGACCATCGAGGAGCTGAGCACCGGATCGTTAAAGAGCTCTCCATGGCC 1512
Db 1157 AGGTG-----ATGCGCCAGCTGGCGCGGCTGACCAACAACTGCTGACGCGC 1207
QY 1513 CGTGCAGACCTGAGGTGCGAGCGGAGCGACAGCGCACCTTTCGAGGAGCTCG 1569
Db 1208 CCAGCGTCAGATGAAGAAATGTCGCGAGGCGCGCATCGAGCGCTGGCGCCCTGG 1264
```

## RESULT 2

```
US-09-252-991A-14144/C
; Sequence 14144, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
```

```
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14144
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14144
```

```
Query Match 4.5%; Score 87.4; DB 4; Length 1338;
Best Local Similarity 46.5%; Pred. No. 2.1e-10;
Matches 585; Conservative 0; Mismatches 636; Indels 36; Gaps 8;
```

```
QY 313 CCATAGCTGTATAGGCTCAGTGATACACACAGCACAGTGGAGATCGGTGAAAAATTG 372
Db 1316 CTTTCAATGCGCTCGGCATCAATACAAAGACCCCTCGGTGGCTGTCGCGAGCGCTGG 1257
QY 373 CTGTTGCAGAGGAACGTGGCCCGTCTATTCAAGAACTCACTAGCCTGAACCATATTG 432
Db 1256 CTTTCACTCCGGAACAGATGGTTCAGGCGCTGCGAGAGCTCTGCCGCTGACCACGACC 1197
QY 433 AAGAGCTGCTGTTCTTAGTACCTGTAATAGAAATGAAATTTATGTGGTGGCGTATCAT 492
Db 1196 GCGAGGCGGCATCTCTGTCGACCTGCNAACGAGCGAATCTTACCTGGAGAT-CGACCAT 1138
QY 493 GGAACCGTGTATCAGAGAGTAGTGAATGCGAAGAAAGTGTATTTCGCGCTT 552
Db 1137 CGAGCGCGGAC-----GACGTGCTGCGCTGGCTGGGACTACCATCGCTGACTCTCG 1083
QY 553 CGAGGCTTAGGAGACCTGTTTCATCTTGGGAAGCAGTGTGCAACGCGCATCTGTTTG 612
Db 1082 ACAGTGTGCGGCTGTGCTATGTGCAACGAGACGCGGTGCGCCACATGATGC 1023
QY 613 AGTGTGTCAGTGGCTTGAATCTTTTGGTCTCGTCAAGAGCAAAATCTTGTCTCAGGTTA 672
Db 1022 CGGTGCGCTCGGCTCGACTCGATGCTCTCGCGAGCGCAGATCTCGCGCAGATGA 963
QY 673 AACAAAGTTGTAGAGTGTGACAGAAACAGTGGAGGCTTGGAAAGAAACATCGATAGATG 732
Db 962 AGTGGCGCTACGCGCTGCGCGCGAGCGCGCACCTGCTGCTGCTGCTGCTGCGCGCTGT 903
QY 733 TCAAGGATCAATCACTGCTGGAAGCGGTGTCGAGCGAGACCAACATATCATCTGTTG 792
Db 902 TCAGGCGCACTTCAGTACGCGGAAGACCGTCCGACCGATACCGCATCGCGGAGAAC 843
QY 793 CTGTTTCTGTCAGTTTCAGCGCGGTTGAACTGCGCCCTGATGAAGCTTCCGAAAGTCTGAAG 852
Db 842 CGGTGTCGCTGGCGTTTCGCGCGGTGAGCTGCGCAAGCAG-ATCTTCAGCGACCTGCAC 784
QY 853 CACTGTCACTAGGATGCTTCTGATTTGGTGTGTTAAATGGGAAGCTAGTATGATCAAC 912
Db 783 CGCAGCCAGGC-----GCTGCTGATCGCGCGCGCGAGACCATCACCTGCTGCGCGGCC 729
QY 913 ATCTGTTGCCAAGAGTGCAGAGGTTGTTGTGTGAACCGCTCGGTGGAAGGTTGG 972
Db 728 ACTGTTTCAGCAAGCGGTGAAGCGCATGCTGTGTGCGCCAAACCGCACCTTGGAGCGGCCA 669
QY 973 ATGCTATTCTGAGGAGATGAAGATATAGAGATCGTGTACAGGCGCTCTCTCAGACATGT 1032
Db 668 GCTGCTCGCGAGCAGTTTCGCGCGCCATGCGGTGCTG-----CTTTCGAGATAC 618
QY 1033 ATCAAGCTGCTGTAAGCTGATGCTGTTTACCAGCAGCAGCGCATCTGAACATTCATGT 1092
Db 617 CGAGGAGTTGGCCAAACAGCATATCGTTCATCAGCTCCACTGCGAGCGAGTTCGCCATCC 558
QY 1093 TCGCAAAAGAACACGAGAGGCACTCCCCCTGCTCTGTATATATATGGAGGTTGTCGCC 1152
Db 557 TCGCAAGGCGCGCTGGAG-CGCGGCTGGAAGAGCGCAAGCAAGCGCGATGTTCA-- 501
QY 1153 TGTTCGTGACATATCTGTCCCGCAGGAATGTTCAGCGCATGTGTGTGTAAGTTGGCGCTG 1212
Db 500 ---TGTTCGACATCGCGCTGCGCGGACATCGAGCGCGGAATGTCGGCGAGCTGGACGAG 444
```

Query Match	4.3%;	Score 82.6;	DB 4;	Length 1026;
Best Local Similarity	49.6%;	Pred. No. 2.2e-09;		
Matches 326;	Conservative 0;	Mismatches 319;	Indels 12;	Gaps 4;

  

Qy	313	CCATAGCTGTGATAGGCCTCAGTGTACACACAGCACCAGTGGAGATCGGTGAAAACTTG	372
Db	141	CCTTTCATTGCCCTCGGCATCAATCACAGACCGCGCTCGGTGGTGTCCGCGAGCGGTGG	200
Qy	373	CTGTTCGAGAGGAACCTGTGCCCGCTGCTATTCAAGAACTCACTAGTGCTCAACCATATTG	432
Db	201	CTTTCACTCGGAACAGATGGTCGAGCGCTCGAGAGCTCTCGCGGCTGCACACAGCC	260
Qy	433	AAGAGGCTGCTGTTCTTTAGTACTGTTAATAGAAATTTATGTGGTGGCGCTATCAT	492
Db	261	GCAGGCGGGCCATCTGTCTGCACCTGCAACCGGAGCGAACTCTACCTGGAGAT-CGACCAT	319
Qy	493	GGAAACGTGCTATCAGAGAGTAGTGACTGGATGTGCAAGAAAGTGGTATTCCCGCTT	552
Db	320	CCGACCGCGAC-----GACGTGCTCGCTGGCTGGCGGACTACCATTGCTGACTCTCG	374
Qy	553	CCGAGCTTAGGGAGCACCTGTTTCATCTTTCGGAAGCAGTGTATGCCACAGCCCATCTGTTTG	612
Db	375	ACGAGTTGCGCGCTGTGGCTATGTGCCACAGACAGAAAGCGGGTCGCGCCACATGATGC	434
Qy	613	AGGTGTACGTGCGCCTTGACTCTTTGGTTCTCGGTGAAGAGCAAAATCTTTGCTCAGGTTA	672

Query Match	3.4%	Score 64.6	DB 4	Length 1347
Best Local Similarity	45.4%	Pred. No. 2.5e-05		
Matches 232	Conservative 0	Mismatches 279	Indels 0	Gaps 0
Qy	316	TAGCTGTGATAGGCTCAGTGTACACACAGCACCGATCGGTGAAACCTTGCTG	375	
Db				
	8	TTGTGGCAATTAGCATAAATCATCGAACACGCTGATGTAAACATTAAAGAGCAAGTTGCTT	67	
Qy	376	TTGCAGAGGAACCTGTGCGCCCGTCTATTCAAGAACTCCTAGCCCTGAACCATATTTGAAG	435	
Db				
	68	TTAGAGATGATGCTTTACGATTAGCCATGAGATTTATATGAAACTAAGCAATTTTAG	127	
Qy	436	AGGCTGCTGTCTTTAGTACCTGTAATAGAAATGGAATTTATGTGTGGCGCTATCATGGA	495	
Db				
	128	AAAATGTCTATTTATCTACATGTAAATCGTACTGAAGTATACGCTATTGTTGATCAAGTTC	187	
Qy	496	ACCGTGTATACAGAGAGTAGTGGACTGGATGTCGAAGAAAGTGTATTCGCGCTTCGG	555	
Db				
	188	ATACAGACGTTATTTATACAAAGATTTTATGCGCGCTCTTTTGGATTTGAGGTAGATG	247	
Qy	556	AGCTTAGGGAGCACCTGTTCACTTTCGAAGCAGTGATGCCACAGCCATCTGTTTGGG	615	
Db				
	248	ATATTAAGATATGTCGGAAGTTAAAGTGGGGACGATGCAGTTGAACATTTATTCGCGT	307	
Qy	616	TGTCAGTGGCCCTTGACTCTTTGGTCTTCGGTGAAGGACAAATCCTTGTCTCAGGTAAAC	675	
Db				
	308	TCACTTCTGGCTTAGATTCAATTTGCTGTGTGAACACAAATTTTATAGGACAAATGCGCG	367	

Qy	676	AGTTGTGAGGAGTGCACAGAACAGTGGAGCTTGGGAAAGAACATCGATAGAGATGTTCA	735
Db	368	ATGCATTTTCTTAGCGCAAAATCTGGTACAACTCGGAACGATTTTAAATCATTTATTTA	427
Qy	736	AGGATCGAATCACTGCTGGAAAGCGTGTCGCGAGCGAGACCAATATCATCTGGTGCTG	795
Db	428	AACAAGCGATTACTTTTGTCTAAAAAGCACACAGTGAACAGACATTTGCAGATAATGCTG	487
Qy	796	TTTCTCTCAGTTTCAGCGCGGTTGAACTGGC	826
Db	488	TGAGTGTGTTCTTATGCTGCTGTTGAATTAGC	518

## RESULT 5

```

US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (101998)..(101998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)

```

```

, NAME/KEY: misc feature
, LOCATION: (871619) .. (871619)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1084930) .. (1084830)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1096846) .. (1096846)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1119881) .. (1119881)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1130881) .. (1130881)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1310988) .. (1310988)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1313224) .. (1313224)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1349473) .. (1349473)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1349491) .. (1349491)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1470091) .. (1470091)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1569020) .. (1569020)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1602912) .. (1602912)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1603734) .. (1603734)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1637998) .. (1637998)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1664854) .. (1664855)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (916-421B-1)
US-08-916-421B-1

```

Query Match	3.3%	Score 64.2;	DB 4;	Length 1664976;
Best Local Similarity	54.4%;	Pred. No. 0.00036;		
Matches 129;	Conservative 0;	Mismatches 108;	Indels 0;	Gaps 0;
QY	590	TGATGCCACACGCCATCTGTTTCGAGGTGTCAGCTGGCCTTGACTCTCTTTGGTTCTCGGTGA	649	
Db	1331928	TAAAGCAATAGACATCTTTTITAGAGTTGCATGTGGTTAGAGTCCATGTTGTTGGAGA	1331869	
QY	650	AGGACAAATCCTTGCTCAGGTTTAAACAGTTGTGAGGATGGGACAGAACAGTGGAGGCTT	709	
Db	1331868	AGACCAAAATACTTGGCGCAGTTAAAAAATGCCTATCTAAAGACAAAAAGAAAGCGCAGAAT	1331809	
QY	710	GGGAAAGAACATCGATAGGATGTTTCAAGGATGCAATCAGCTGCTGGAAAGCGTCTCCGCAG	769	
Db	1331808	ATCCAAAAAATTCGAGAAAATTAATTTTAAAGGCAATACATCTGGACAAAGGCGCAGAGT	1331749	
QY	770	CGAGACCAACATATCATCTGGTGCTGTTTCTGTGACTTTCAGCGCGGCTGAACTGGC	826	
Db	1331748	AGAGACAAAGATAAATGAGCGCGGGGTTTCAATTGGCTCTCGCGCAGTTGAAATTGGC	1331692	

RESULT 6  
US-09-328-352-517  
; Sequence 517, Application US/09328352  
; Patent NO. 6562958  
; GENERAL INFORMATION:

```

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 517
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-517

```

Query Match	3.0%;	Score 57.4;	DB 4;	Length 1290;
Best Local Similarity	45.2%;	Pred. No. 0.001;		
Matches 309;	Conservative 0;	Mismatches 356;	Indels 18;	Gaps 2
QY	320	TGFGATAGGCCTCAGTGTACACAGCAGCAGTCGGATGGTGAAAACTCGCTGTGC	3799	
Db	18	TGCATTTGGGTCAACCATCAACAGCTTCTGTAGAACTCCGCGAGCAAAATGCTTTCAA	77	
QY	380	AGAGGAACTGTGGCCCGCTGCTATTTCGAAGAACTCACTAGCTGAACCAATATTGAAGAGGC	4399	
Db	78	TCGAGAGCGATTAAAGTAAATTTGCTTGGCGAACACGTCATCAGAGTCTTTAAAGACTT	1379	
QY	440	TGCTGTTCTTAGTACCTGTAAATAGAAATGGAAATTTATGTGTGGCGCTATCATGGAACCG	4999	
Db	138	AGTGGTTGTCTCAACGTTAAACCGTACAGAACTCTACGCAATGGCGAA-----	1869	
QY	500	TGGTATCAGAGAGTAGTGGACTGGATGTCGAAGAAAGTGGTATTCCCGCTTCCGAGCT	5599	
Db	187	-CATGCCGAAAGCCCTTCTGAAATGGTGTAGTGTGCCAAATAATTGATGTAAGCAGTT	2459	
QY	560	TAGGAGGACACTGTTTCATCTTTCGAAGCAGTGTGCCACAGCCATCTGTTTGCAGGTGTC	6199	
Db	246	AAITTCATCATGTTTATCGTTACGAAATGCTCAAGCCATCACACATTTAATGCGTGTAGC	3059	
QY	620	AGCTGGCCTTGACTCTTTGGTTCTCGGTGAAGCAAAATCCTTGCTCAGGTTAAACAAGT	6799	
Db	306	AAATGGTTTGGACTCACTCATGTTGGGCGAACCCGAAATTTTAGGGCAGGTGAAAAGTGC	3659	
QY	680	TGTGAGGAGTGGACAGAACAGTGCAGGCTTGGGAAAGAACATCGATAGAGTGTTCAGGA	7399	
Db	366	TCACATTTGCTTAAAGAGCGCAACAGTTTCTCCAGAAATTAATAGTGTATTTCAGTA	4259	
QY	740	TGCAATCACTGCTGGAAGCGTGTCCGAGCGAGCACCAACATATCATCTGGTGTGTTTTTC	7999	
Db	426	CGCTTTTACCGCGCTAAACGTTGCGCTCTGAAACAGCTGTAGGTAGTCATCGCGTTTC	4859	
QY	800	TGTCAGTTACGGCGGTTGAACCTGGCCCTGTATGAAGCTTCCGAAGTCTGAAGCACTGTC	8599	
Db	486	AATGGGTTTATGCAGTTGCCAGCTCGCTTTACAGGTTTTTTAGTAAACCTGAAAAAATAAC	5459	
QY	860	AGCTAGGATGCTTCTGATTTGGTGTCTGTGTAAATGGGAAAGCTAGTGNATCAAACTCTGGT	9199	
Db	546	GGTTA-----TGGTGGTTGCTGCTGGCGAAATGAACAGCCTTTGTTGCCAAGCATTTAGC	5999	
QY	920	TGCCAAAGGATGCAAGAGGTTGTTGTGGTGAACCCGCTCCGTGGAAAGGGTGTATGCTAT	9799	
Db	600	TGAATTTGGCGCTGCCAAGATGATTATTTCGTATCTGTAGCCGTGAACGTCGGATCAATT	6599	
QY	980	TGCTGAGGAGATGAAGATATAG 1002		
Db	660	GGCCCAAGAAATTCACATCAAG 682		

RESULT 7  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.



APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 2.8%; Score 53; DB 3; Length 4411529;  
Best Local Similarity 45.1%; Pred. No. 0.16;  
Matches 243; Conservative 0; Mismatches 290; Indels 6; Gaps 1;

QY 435 GAGGCTGCTGTTCTTAGTACCTGTAATAGATGGAATTTATGTTGGTGGCGCTATCATGG 494  
DB 600565 GAGGCGATGGTCTGTGACCTTGCAACCGCTGAGGTCTACGCCGTAGTGACCGGTTC 600624  
QY 495 AACCGTGTATCAGAGAAGTAGTGACTGGATGTCGAAGAAAGTGGTATTCCCGCTTCC 554  
DB 600625 CATGCGCCCTGTGCGTGTATCGGCGAGGTGCTTCCGAACACTCCGGTATGTCGATGGG 600684  
QY 555 GAGCTTAGGGAGCACCTGTTTATCTTGGAGCACTGATGCCACAGCCATCTGTTTGA 614  
DB 600685 GAGCTGACCAAGTACGCATATGTCCGCTACAGCAGGCGAGCTTGAGCACCTGTTCGCG 600744  
QY 615 GTGTCAGCTGGCTTGACTTTTGGTCTCCGTGAAGCAGCAATCTTCTCAGGTTAA 674  
DB 600745 GTTCCACGCGCTGGACTCGCGGTGATCGCGAGCAGAGTCTTGGTCAGGTGCGC 600804  
QY 675 CAAGTTGTAGGAGTGGACAGAACAGTGGAGCTTTGGAAAGAACATCGATAGGATGTC 734  
DB 600805 CGCGCTATGCGTGCAGCAATCCAAACCGCAGGTGCGCGGTGTCACGAATTGGCC 600864  
QY 735 AAGATGCAATCACTGTGGAAGCGTTCGCGAGCAGACCAATCATCTGTGTTGCT 794  
DB 600865 CAGCGGGCGTGTGCGTGGGCAAGCGAGTGCACCTCCGAACCGCCATTTGACGCTGCCGT 600924  
QY 795 GTTCTGTGAGTTCAGCGCGGTGTAATCTGCGCTGATGAAGCTTCGAACTCTGAACCA 854  
DB 600925 GCCTCGTGGTGTGCTGCTGCTGCTGGAATGCGCGAGCCCAATTGGGCTGTTGGCGGC 600984  
QY 855 CTGTCAGTAGGATGCTTCTGATTGCTGCTGTAATGGAAAGCTAGTATCAACAT 914  
DB 600985 ACGACCGC-----GGTGTGATCGCGCGCGGCGATGGCGCGCTGTGCGGTACAT 601038  
QY 915 CTGTTGCAAGGATGCAAGAGTGTGTTGTTGTAACCGCTCGTGGAAAGGTTGA 973  
DB 601039 CTGACCCGTGCGCGGTGCGGCACATTCAGGTGCTCAACCGTGTGTTGCTCCCGCGCA 601097

RESULT 10  
US-09-514-247A-9/c  
Sequence 9, Application US/09514247A  
Patent No. 6365361  
GENERAL INFORMATION:  
APPLICANT: TANABE SEIVAKU CO. LTD.  
APPLICANT: TANIGUCHI, Tomoyasu  
APPLICANT: MIZUKAMI, Junko  
TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPA  
FILE REFERENCE: TANIGUCHI=6  
CURRENT APPLICATION NUMBER: US/09/514,247A  
CURRENT FILING DATE: 2000-02-28  
PRIOR APPLICATION NUMBER: PCT/JP98/03734  
PRIOR FILING DATE: 1998-08-24  
PRIOR APPLICATION NUMBER: JP231084/1997  
PRIOR FILING DATE: 1997-08-27  
NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn version 3.0  
SEQ ID NO 9  
LENGTH: 8147  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (819)..(8147)  
US-09-514-247A-9

Query Match 2.6%; Score 50; DB 4; Length 8147;  
Best Local Similarity 55.1%; Pred. No. 0.086;  
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 82 CGACGTCTAGCAGCACACCGCGCGGCGGAGCAGCGCGCCACCGCCAAAGCGCGCGGGGTGCT 141  
DB 295 CGCGAGCGGGGTGCG 236  
QY 142 CGTGGCCCTCTGCCAGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 201  
DB 235 CGCGCTGCG 176  
QY 202 CGCACG 259  
DB 175 CG 118

RESULT 11  
US-08-311-731A-128/c  
Sequence 128, Application US/08311731A  
Patent No. 6583266  
GENERAL INFORMATION:  
APPLICANT: SMITH, DOUGLAS  
APPLICANT: MAO, JEN-I  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 411  
CORRESPONDENCE ADDRESS:  
ADDRESS: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,731A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: C0044/7125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
INFORMATION FOR SEQ ID NO: 128:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42988 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: MYCOBACTERIUM LEPRAE  
US-08-311-731A-128

Query Match 2.6%; Score 49.6; DB 4; Length 42988;  
Best Local Similarity 45.4%; Pred. No. 0.19;  
Matches 178; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 435 GAGGCTGCTGTTCTTAGTACCTGTAATAGAAATGGAATTTATGTGTCGCGCTATCATGG 494  
DB 20252 GAGGCGATGGTGCTGTGACCTGTAACTGTCGAGGTTTACCGCGTGGTGAACGGTTC 20193

QY 495 AACCGTGGTATCAGAGAAGTAGTGACTGGATGTCGAAGAAAGTGGTATTCCTCCGCTTCC 554  
DB 20192 CACGCGGATTATCAGTGATTGGCAAGTGTGCGGAATACTCCGCGATGTCGATCGGG 20133

QY 555 GAGCTTAGGAGCACCTGTTCTATCTTGGGAACAGTGATGCCACAGCCATCTGTTTGG 614  
DB 20132 GACCTGACCAATAACGCTTATGACCTACACGAGGCTGTGTCGAGCATTGTTACG 20073

QY 615 GTGTCAGCTGGCGCTTGACTCTTTGTTCTCGGTGAAGCAAAATCTTCTCAGGTTAA 674  
DB 20072 GTTGCCAGCGTCTGGATTACGGTGTGTCGCGAGCAGCAGTGTGTTGTTGAGTGGT 20013

QY 675 CAAGTTGTGAGAGTGGACAGAACAGTGAGGCTTGGGAAGAAACATGATAGGATGTC 734  
DB 20012 CGCGGCTATGCGCGCGCTGAGGCTAATCGCACGCTAGGCCAGGTGTGTCATGAGTGGCC 19953

QY 735 AAGGATGAATCACTGCTGGAAGCGTGTCCGACGAGACCAATATCATCTGTTGCT 794  
DB 19952 CAACGAGCGTTGTCGTTGAGGTTGAAGCGGTCATTCGAAACTGCCATTGACCTGCCGG 19893

QY 795 GTTCTGTGTCAGTTCAGCGCGCTTGAACCTGGC 826  
DB 19892 GTCTCGTGGTGTGCTGCGTCCGCTTGGCATAGC 19861

RESULT 12  
US-09-086-010-1/c  
; Sequence 1, Application US/09086010  
; Patent No. 6274338  
; GENERAL INFORMATION:  
; APPLICANT: Glimcher, Laurie H. et al.  
; TITLE OF INVENTION: Human c-Maf Compositions and  
; TITLE OF INVENTION: Methods of Use Thereof  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: LAHIVE & COCKFIELD, LLP  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/086,010  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/030,579  
; FILING DATE: 2-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kara, Catherine J.  
; REGISTRATION NUMBER: 41,106  
; REFERENCE/DOCKET NUMBER: HUI-027CP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)742-4214  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1203 base pairs  
; TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1203  
US-09-086-010-1

Query Match 2.6%; Score 49.4; DB 3; Length 1203;  
Best Local Similarity 51.6%; Pred. No. 0.061;  
Matches 113; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 22 GAGAGCTTGGGAAGTTCCTCCAAAGGCTTCGATTTTCGCTAATGCGACGATAATGGCGACCA 81  
DB 776 GAGAAGCGGTGCTGCGAAGTGCAGGCGCGCGCGGTGCGGTGCGAGGCGCGCGCC 717

QY 82 CGAGCTCAGCGACCAACCGCCCGCGCAGCAGCGCCACCGCCCAAGCGCGCGCGGTCT 141  
DB 716 GCGCCCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657

QY 142 CGTCGCGCTCTGCGCAGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 201  
DB 656 CG 597

QY 202 GCGAGCGCGCGCGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGT 240  
DB 596 GCGCGCGCGGTGCGGTGCGGTGCGCGCGCGCGCGCGCGT 558

RESULT 13  
US-08-466-103A-3  
; Sequence 3, Application US/08466103A  
; Patent No. 5856124  
; GENERAL INFORMATION:  
; APPLICANT: Reppert, Steven M.  
; APPLICANT: Ebisawa, Takashi  
; TITLE OF INVENTION: HIGH-AFFINITY MELATONIN  
; TITLE OF INVENTION: RECEPTORS AND USES THEREOF  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,103A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/319,887  
; FILING DATE: 07-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/261,857  
; FILING DATE: 17-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frazer, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 00786/250002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1149 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double





GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 12:53:43 ; Search time 39 seconds  
(without alignments)  
3513.485 Million cell updates/sec

Title: US-10-018-902-4  
Perfect score: 2606  
Sequence: 1 MATTTTATTAATAAATTAKP.....DWEKALIEQIKAKVEKTON 531

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 23:\*

1:	sp_archaea:*
2:	sp_bacteria:*
3:	sp_fungi:*
4:	sp_human:*
5:	sp_invertebrate:*
6:	sp_mammal:*
7:	sp_mhc:*
8:	sp_organelle:*
9:	sp_page:*
10:	sp_plant:*
11:	sp_rodent:*
12:	sp_virus:*
13:	sp_vertebrate:*
14:	sp_unclassified:*
15:	sp_virus:*
16:	sp_bacteriap:*
17:	sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2393.5	91.8	536	10 Q9FW00	Q9fw00 oryza sativ
2	1789	68.6	542	10 Q9ZPK4	Q9zpk4 glycine max
3	1440	55.3	339	10 Q9FVS1	Q9fyv1 hordeum vul
4	1334	51.2	524	10 Q9SUX1	Q9sux1 arabidopsis
5	1326.5	50.9	522	10 Q9FPR7	Q9fpr7 chlamydomon
6	1063	40.8	252	10 Q9FVS2	Q9fyv2 hordeum vul
7	910.5	34.9	248	10 Q9AVP6	Q9avp6 amaranthus
8	893.5	34.3	426	16 Q8DI53	Q8di53 synechococc
9	646	24.8	419	16 Q8DFF8	Q8dff8 vibrio vuln
10	618.5	23.7	446	2 Q8GCB0	Q8gcb0 bacillus me
11	597	22.9	416	16 Q8EAR2	Q8ear2 shewanella
12	588	22.6	457	16 Q8EPM7	Q8epm7 oceanobacil
13	570.5	21.9	484	16 Q8FT03	Q8ft03 escherichia
14	543.5	20.9	448	16 Q8CNY6	Q8cny6 staphylococ
15	488.5	18.7	463	2 Q8GMZ2	Q8gmz2 corynebacte
16	483.5	18.6	463	16 Q8NT91	Q8nt91 corynebacte

17	465.5	17.9	466	16 Q8FSP8	Q8fsf8 corynebacte
18	451	17.3	460	17 Q8TT60	Q8tt60 methanosarc
19	421	16.2	420	16 Q8D2K8	Q8d2k8 wigglewort
20	294	11.3	329	16 Q8R687	Q8r687 fusobacteri
21	287	11.0	56	10 Q8RUB4	Q8rub4 zea mays (m
22	282	10.8	55	10 Q8S2X3	Q8s2x3 zea mays (m
23	277	10.6	36	10 Q8LNE9	Q8lne9 oryza sativ
24	256.5	9.8	377	2 Q9JA66	Q9ja66 uncultured
25	190	7.3	184	2 Q9AQ47	Q9aq47 selenomonas
26	156.5	6.0	644	16 Q8RAZ5	Q8raz5 thermomanae
27	152	5.8	807	10 Q8H724	Q8h724 arabidopsis
28	152	5.8	816	5 Q8IU20	Q8iu20 branchiosto
29	150	5.8	794	17 Q8ZVH4	Q8zvh4 pyrobaculum
30	148	5.7	853	10 Q9LFE4	Q9lfe4 arabidopsis
31	147.5	5.7	739	17 Q8TUP6	Q8tup6 methanosarc
32	147.5	5.7	791	17 Q9YB89	Q9ybh9 aeropyrum p
33	147.5	5.7	1307	16 Q8K792	Q8k792 streptococc
34	147	5.6	530	16 Q9XOM7	Q9xom7 thermotoga
35	144.5	5.5	1751	5 Q8I0G9	Q8i0g9 plasmodium
36	143.5	5.5	599	2 Q44107	Q44107 anaplasmia m
37	143.5	5.5	779	10 Q9SZB6	Q9szb6 arabidopsis
38	143.5	5.5	817	10 Q8LHQ6	Q8lhq6 oryza sativ
39	143.5	5.5	1307	16 Q8P1P0	Q8plp0 streptococc
40	142	5.4	2158	10 Q9LUT5	Q9lut5 arabidopsis
41	140.5	5.4	1313	10 Q9XIP6	Q9xip6 arabidopsis
42	140	5.4	1963	5 Q02244	Q02244 caenorhabdi
43	139.5	5.4	1728	10 Q9LUI2	Q9lui2 arabidopsis
44	139	5.3	1300	4 Q94986	Q94986 homo sapien
45	139	5.3	1978	4 Q15154	Q15154 homo sapien

## ALIGNMENTS

## RESULT 1

Q9FW00 PRELIMINARY; PRT; 536 AA.

AC Q9FW00; TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Putative glutamyl-tRNA reductase (EC 1.2.1.-) (GLUTR).

GN OSUNB0073N24.1.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Eriactoidae; Oryzeae; Oryza.

OX NCBI\_taxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;

RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,

RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,

RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblum T.V.,

RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;

RT "Oryza sativa chromosome 10 BAC OSUNB0073N24 genomic sequence.";

RL submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: GLUTAMYL-TRNA (GLU) + NADPH = GLUTAMATE-1- SEMIALDEHYDE + NADP(+) + TRNA (GLU).

CC -1- PATHWAY: PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY, FIRST STEP.

CC -1- SIMILARITY: BELONGS TO THE GLUTAMYL-TRNA REDUCTASE FAMILY.

DR EMBL; AC078840; AAG13620.1; -.

DR Gramene; AC078840; -.

DR InterPro; IPR000343; Glutr.

DR Pfam; PF00745; Glutr; 1.

DR TIGRfam; TIGR01035; hema; 1.

DR PROSITE; PS00747; Glutr; 1.

KW NADP; Oxidoreductase; Porphyrin biosynthesis.

SO SEQUENCE 536 AA; 58288 MW; 4C1C7E4A58E3C640 CRC64;

Query Match 91.8%; Score 2393.5; DB 10; Length 536;  
Best Local Similarity 91.6%; Pred.No. 3.1e-132;  
Matches 492; Conservative 15; Mismatches 23; Indels 7; Gaps 4;

QY 1 MATTTTATATAAAATTAATKPRGSSA---LQORVAGGRRRSG-VVRCDAAGVE--AAQA 54  
DB 1 MASTTSAAGGAFAAATKTRAGSSAAGGACARVAAGGRRRSGVVVRCD-AGVEAAQAQ 59  
QY 55 AVAKAASVAALAEQFKISADRYMKERSTTAVIGLSVHTAPVEMREKLAVALWPAIOEL 114  
DB 60 AVAKAASVAALAEQFKISADRYMKERSTTAVIGLSVHTAPVEMREKLAVALWPAIOEL 119  
QY 115 TSLNHIIEAAVLSTCNRMIEIYVVALSWNRGIREVVDMWMSKSGIPASBLRHLFIRSSD 174  
DB 120 TSLNHIIEAAVLSTCNRMIEIYVVALSWNRGIREVVDMWMSKSGIPASBLRHLFIRSSD 179  
QY 175 ATRHLFEVSAGLDSVLGEGQILAQVKQVVRSGQSGGLGNIDRMFKDAITAGKRVSE 234  
DB 180 ATRHLFEVSAGLDSVLGEGQILAQVKQVVRSGQSGGLGNIDRMFKDAITAGKRVSE 239  
QY 235 TNISSGAVSVSSAAVELALMKLPKSEALSARMMLIGAGKMGKLVKHLVAKGCKVWVYN 294  
DB 240 TNISSGAVSVSSAAVELALMKLPKSEALSARMMLIGAGKMGKLVKHLVAKGCKVWVYN 299  
QY 295 RSVERDAIREEMKDIETVYRPLSDMYQAAAEADVFTSTASSETSLFAKEHAELPPVSD 354  
DB 300 RSVERDAIREEMKDIETVYRPLSDMYQAAAEADVFTSTASSETSLFAKEHAELPPVSD 359  
QY 355 TMGGVRLFVDISVPRNVACVSEVGAARVYVDDLKEVVEANKEDRLKAMEAQTIITEE 414  
DB 360 AMGGVRLFVDISVPRNVACVSEVGAARVYVDDLKEVVEANKEDRLKAMEAQTIITEE 419  
QY 415 LRRFEAWRDSLETPTIKKLSYADRIASELEKCLQKVGEDALTKKMRRAIEELSTGIV 474  
DB 420 LRRFEAWRDSLETPTIKKLSYADRIASELEKCLQKVGEDALTKKMRRAIEELSTGIV 479  
QY 475 NKLHGPIQLHRCDSGSDSRTLDTELNNHMLNRMFSLDMEKAIIEQIKAKVKTQN 531  
DB 480 NKLHGPIQLHRCDSGSDSRTLDTELNNHMLNRMFSLDMEKAIIEQIKAKVKTQN 536

RESULT 2  
Q9ZPK4 PRELIMINARY; PRT; 542 AA.  
ID Q9ZPK4  
AC Q9ZPK4;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-MAY-2003 (TREMBlrel. 23, Last annotation update)  
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (Glutr).  
GN GTR1.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN=Essex;  
RX MEDLINE=99137912; PubMed=9952455;  
RA Sangwan I., O'Brian M.R.;  
RT "Expression of a soybean gene encoding the tetrapyrrole-synthesis  
enzyme glutamyl-tRNA reductase in symbiotic root nodules.";  
RU Plant Physiol. 119:593-598(1999).  
CC -1- CATALYTIC ACTIVITY: GLUTAMYL-TRNA (GLU) + NADPH = GLUTAMATE-1-  
SEMIALDEHYDE + NADP(+) + TRNA (GLU).  
CC -1- PATHWAY: PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY, FIRST STEP.  
CC EMBL; AF105221; AAD16897.1; -.  
DR InterPro; IPR000343; Glutr.  
DR InterPro; IPR000594; Thif domain.  
DR Pfam; PF00745; Glutr.1.  
DR TIGRFAMs; TIGR01035; hema; 1.  
DR PROSITE; PS00747; Glutr.1.  
DR NADP; Oxidoreductase; porphyrin biosynthesis.  
SQ SEQUENCE 542 AA; 59204 MW; CB85AA5494F09792 CRC64;

Query Match 68.6%; Score 1789; DB 10; Length 542;  
Best Local Similarity 67.3%; Pred. No. 8e-97;  
Matches 366; Conservative 70; Mismatches 88; Indels 20; Gaps 5;  
QY 1 MATTTT-----ATTAATAAATTAATKPRGSSSALCORVAGGRRRSGVVVRCD 46  
DB 1 MAVSTSPGAKLEALLKCGSSNAATATATTTT---HLSCFKTRKTLVQSGRGPIRCEA 57  
QY 47 AGV-BAQAQAVAKAASVAALAEQFKIS-ADRYMKERSTTAVIGLSVHTAPVEMREKLAVAL 104  
DB 58 SSASDVVADATKKAASVSALAEQFKISADRYTKERSVVMVIGLSVHTSPVEMREKLAIP 117  
QY 105 ELMPRATQELTSLNHIIEAAVLSTCNRMIEIYVVALSWNRGIREVVDMWMSKSGIPASBLR 164  
DB 118 AEMPRATAECLSLNHIIEAAVLSTCNRMIEIYVVALSKRGVKEVTEMNMTSGIPVADLC 177  
QY 165 EHLFILRSSDATRELFVFSAGLDSVLGEGQILAQVKQVVRSGQSGGLGNIDRMFKDA 224  
DB 178 QHQFLLYNKDATQHLFEVSAGLDSVLGEGQILAQVKQVVGQVNGFGRNISGLFKHA 237  
QY 225 ITAGKRVSRSETNISGAVSVSSAAVELALMKLPKSEALSARMMLIGAGKMGKLVKHLVA 284  
DB 238 ITVGKRVRTETNIAAGAVSVSSAAVELALMKLPKSEASHANARMLVIGAGKMGKLVKHLVA 297  
QY 285 KGCKVWVVRNRSVERDAIREEMKDIETVYRPLSDMYQAAAEADVFTSTASSETSLPAKE 344  
DB 298 KGCTMNVVVRNRSVERDAIREEMKDIETVYRPLSDMYQAAAEADVFTSTASSETSLPAKE 357  
QY 345 HAEALPPVSDTMGGVRLFVDISVPRNVACVSEVGAARVYVDDLKEVVEANKEDRLKRA 404  
DB 358 DVKELPATDEVGGRRLFVDISVPRNVACVSEVGAARVYVDDLKEVVEANKEDRLKRA 417  
QY 405 MEAQTIITEELRRFEAWRDSLETPTIKKLSYADRIASELEKCLQKVGEDALTKKMR 464  
DB 418 MEAQTIIEELRRFEAWRDSLETPTIKKLSYADRIASELEKCLQKVGEDALTKKMR 476  
QY 465 AIELSTGIVNKLHGPIQLHRCDSGSDSRTLDTELNNHMLNRMFSLDMEKAIIEQIKAK 524  
DB 477 AVDDLRSRGIVNKLHGPIQLHRCDSGSDSRTLDTELNNHMLNRMFSLDMEKAIIEQIKAK 536  
QY 525 KVEK 528  
DB 537 KVEQ 540

RESULT 3  
Q9FY51 PRELIMINARY; PRT; 339 AA.  
ID Q9FY51  
AC Q9FY51;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (Glutr) (Fragment).  
GN HEMA2.  
OS Hordeum vulgare (Barley).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
OC Triticeae; Hordeum.  
OX NCBI\_TaxID=4513;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN=cv. Bonus;  
RA Hanson M.;  
RT "Basic characterization of two barley hema promoter regions.";  
RU Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: GLUTAMYL-TRNA (GLU) + NADPH = GLUTAMATE-1-  
SEMIALDEHYDE + NADP(+) + TRNA (GLU).  
CC -1- PATHWAY: PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY, FIRST STEP.  
CC -1- SIMILARITY: BELONGS TO THE GLUTAMYL-TRNA REDUCTASE FAMILY.  
DR EMBL; AF294753; AAG02480.1; -.  
DR InterPro; IPR000343; Glutr.  
DR Pfam; PF00745; Glutr.1.  
DR PROSITE; PS00747; Glutr.1.



```
Query Match      50.9%; Score 1326.5; DB 10; Length 522;
Best Local Similarity 53.7%; Pred. No. 9.4e-70;
Matches 275; Conservative 79; Mismatches 121; Indels 37; Gaps 5;

Qy 2 ATTTTATTAATAAATTAATKPRGSSSALCORVAGGRRSGVVRCDAAAGVAAQAQAA 61
Db 34 AATQTATTAASSTTKLP-----ASHLE-----SS 59

Qy 62 VAALAEQFKISA-DRYM-KERSTIAVIGLSVHTAPVEMREKLAVALPRAIOELTSLNH 119
Db 60 KKALDSLKQAVNRYAGDKSSITAGLTINAPVELREKLAVALPRAIEELCOFPH 119

Qy 120 IEAAVLSTCNRMELIYVVALSWNRGIREVDVDMKSGIPASELREHLFILRSSDATHL 179
Db 120 IEAAVLSTCNRMELIYVVALSWNRGIREVDVDMKSGIPASELREHLFILRSSDATHL 179

Qy 180 FEVSAGLDSLVLGEGQILAQVQVVRSGNSGGGLGKNDRMFKDAITAGKVRSETNIS 239
Db 180 MRVSGGLDSLVMGEGQILAQVQVVRSGNSGGGLGKNDRMFKDAITAGKVRSETNIS 239

Qy 240 GAVSVSSAAVELALMKLPKSPALSGARMILLIAGAGKMGKLVIRKLVAKGCKKYVVNRSVER 299
Db 240 GSVSVSSAAVELALMKLPKSPALSGARMILLIAGAGKMGKLVIRKLVAKGCKKYVVNRSVER 299

Qy 300 VDAIREENKDJIEIYRPLSDMYQAAAADVVFTTASSETSLFAKEHAELPPVSDTMGGV 359
Db 300 AQAALAEPEVKFNIHLMPDLLOQVEASDVIFAASGSEELIHKHVEAMSKPSDVVVGSK 359

Qy 360 RLFDVDSVPRNVSAVSVGAARVYVVDLKEVVEANKEDRLRKAMEAQTIITELRFE 419
Db 360 RRFVDSVPRNIAINELEHGIYVVDLKEVVAANKEGRAQAAAEVLIRBEOQAFE 419

Qy 420 AWRDSLETVPITIKLRSYADRIRASELEKICLOKVGEDALTCKMRAIEELSTGVNKLH 479
Db 420 AWRDSLETVPITIKALRSKAETIRAAEPEKAVSRIGE-GLSKQLKAVELSKGVNKLH 478

Qy 480 GPLQHLRCDGSDSLTDELTLENMHALNRMFSL 511
Db 479 GPMTALRCDDGTPDAVGQTLNMEALERMFO 510

RESULT 6
Q9FY52 PRELIMINARY; PRT; 252 AA.
AC Q9FY52
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR) (Fragment).
GN HEMAL.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bonus;
RA Hansson M.;
RT "Basic characterization of two barley hema promoter regions.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: GLUTAMYL-TRNA(GLU) + NADPH = GLUTAMATE-1-
CC SEMIALDEHYDE + NADP(+) + TRNA(GLU).
CC -1- PATHWAY: PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY; FIRST STEP.
CC DR EMBL; AF294752; AAG02479.1; -.
CC InterPro; IPR000343; GluTR.
DR Pfam; PF00745; GluTR; 1.
DR PROSITE; PS00747; GluTR; 1.
KW NADP; Oxidoreductase; Porphyrin biosynthesis.
FT NON TER 252
FT SEQUENCE 252 AA; 26654 MW; 4D3C0908EAAAF0FE CRC64;
```

```
Query Match      40.8%; Score 1063; DB 10; Length 252;
Best Local Similarity 86.7%; Pred. No. 9.1e-55;
Matches 222; Conservative 10; Mismatches 20; Indels 4; Gaps 2;

Qy 1 MATTTTATTAATAAATTAATKPRGSSSALCORVAGGRRSGVVRCDAAAGVAAQAQAA 60
Db 1 MAGATSA-TAAGAFAAAKARGPAAACPMWVAAGRRRSGVVRCDAGG--DAQAASKAA 56

Qy 61 SVAALAEQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALPRAIOELTSLNH 120
Db 57 SITALEQFKISADRYMKERSSIAVIGLSVHTAPVEMREKLAVALPRAISELTSLNHI 116

Qy 121 IEAAVLSTCNRMELIYVVALSWNRGIREVDVDMKSGIPASELREHLFILRSSDATHL 180
Db 117 IEAAVLSTCNRMELIYVVALSWNRGIREVDVDMKSGIPASELREHLFILRSSDATHL 176

Qy 181 FEVSAGLDSLVLGEGQILAQVQVVRSGNSGGGLGKNDRMFKDAITAGKVRSETNIS 240
Db 177 EVSAGLDSLVLGEGQILAQVQVVRSGNSGGGLGKNDRMFKDAITAGKVRSETNIS 236

Qy 241 AVSVSSAAVELALMKL 256
Db 237 AVSVSSAAVELAMMKL 252

RESULT 7
Q9AVF6 PRELIMINARY; PRT; 248 AA.
AC Q9AVF6
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Glutamyl-tRNA reductase (Fragment).
GN GLUTR.
OS Amaranthus tricolor (Tampala).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Amaranthaceae; Amaranthus.
OX NCBI_TaxID=29722;
RN [1]
RP SEQUENCE FROM N.A.
RA Iwamoto K., Fukuda H., Sugiyama M.;
RT "Elimination of POR Expression correlates with Red Leaf Formation in
RT Amaranthus tricolor.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050119; BAB41186.1; -.
DR InterPro; IPR000343; GluTR.
DR Pfam; PF00745; GluTR; 1.
FT NON TER 1
FT NON TER 248
FT SEQUENCE 248 AA; 27814 MW; 99C31D03F7B4D4BB CRC64;

Query Match      34.9%; Score 910.5; DB 10; Length 248;
Best Local Similarity 72.5%; Pred. No. 7.6e-46;
Matches 177; Conservative 31; Mismatches 35; Indels 1; Gaps 1;

Qy 269 IGAGKMGKLVIKHLVAKGCKKYVVVNRVSDAIREEMKDIEIVYRPLSDMYQAAAEAD 328
Db 1 IGAGKMGKLVIKHLVAKGCKKYVVVNRVSDAIREEMKDIEIVYRPLSDMYQAAAEAD 328

Qy 329 VFTSTASSTSLFAKEHABALPPVSDTMGGVRLFVDSVPRNVSAVSVGAARVYVVD 388
Db 61 VFTSTASSTSLFAKEHABALPPVSDTMGGVRLFVDSVPRNVSAVSVGAARVYVVD 388

Qy 389 LKEVVAANKEDRLRKAMEAQTIITELRFEAWRDSLETVPITIKLRSYADRIRASELEK 448
Db 121 LKEVVAANKEDRLRKAMEAQTIITELRFEAWRDSLETVPITIKLRSYADRIRASELEK 448

Qy 449 CLKQVGEDALTCKMRAIEELSTGVNKLHGLQHLRCDGSDSLTDELTLENMHALNRM 508
Db 181 CLKQVGEDALTCKMRAIEELSTGVNKLHGLQHLRCDGSDSLTDELTLENMHALNRM 508
```

```

QY 509 FSLD 512
DB 240 FSLE 243

RESULT 8
QBDI53
ID QBDI53 PRELIMINARY; PRT; 426 AA.
AC QBDI53;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Transfer RNA-Gln reductase.
GN HWA OR TLL1738.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005375; BAC09290.1; -.
KW Complete proteome.
SQ SEQUENCE 426 AA; 47596 MW; D84CESA1D2AA77E CRC64;

Query Match 34.3%; Score 893.5; DB 16; Length 426;
Best Local Similarity 44.2%; Pred. No. 1.6e-44;
Matches 193; Conservative 87; Mismatches 134; Indels 23; Gaps 5;

QY 82 IAVIGLSVHTAPVEMREKLAVALAEALPRAIOELTSLNHIIEAAVLSTCNRMETIVVALSW 141
DB 1 IAVIGLSVHTAPVEMREKLAVALAEALPRAIOELTSLNHIIEAAVLSTCNRMETIVVALSW 141
DB 3 IAVIGLSVHTAPVEMREKLAVALAEALPRAIOELTSLNHIIEAAVLSTCNRMETIVVALSW 141
QY 142 NRGIREVVDWMSKSGIPASELREHLFILRSSDATHLFEVSAGLSLVLGEGOILAQVK 201
DB 63 EVGVREVHQFLSEWSHPLQPLPYLIFILLHQAAMHLMRVASGLDSLVLGEGOILSQVK 122
QY 202 QVRSQNGSGGLGNIDRMFKDAITAGKRVSETNISGAVSVSSAAVELAMKLPSEA 261
DB 123 RCHOLGOOYKSGIPILNRVFTGATAAGKRVETSTGTAAGVSVSSAAVELADRLQLQ- 181
QY 262 LSARMLLIGAGKMGKLVKHLVAGCKVWVNVRSVERVDAIREEMKDIEIVRPLSDMY 321
DB 182 -NCRIAVVAGKMGKLVVQHLIARGVKEIRIINSLERAQEAQOFPFELFTMTDLL 240
QY 322 QAAAEADVFTSTASLFLAKHAEALPPVSDTMGGVRL-----FVDISVPRNSACV 375
DB 241 PIVAAADLVFTSTAATEPLDR-----DNLGAVLVGDRSLAIDISVPRNVHANV 290
QY 376 SEVGAARYVVDLKEVVEANKEDRLRKAMEAQIITEELRRFEAWRDSLETPTIKKL 435
DB 291 TELGTQVLFNVDDLQAVVAQNEARQAQEAEGILEELETFLAWHALETVPRIISLR 350
QY 436 SYADIRIASELEKLCQKVGEDALTKMRRATEELSTGIVNKLKPLQLHRCDSRSL 495
DB 351 QKMEAIRTOLEKALSLRGSF-PADKQGVIEAMTRTIINKILHDPVTQVQ-----SORDL 405
QY 496 DETLENHMLNRMSLD 512
DB 406 ESRQAMOTLQDLFNL 422

RESULT 9
QBDFF8
ID QBDFF8 PRELIMINARY; PRT; 419 AA.
AC QBDFF8;

```

```

DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Glutamyl-tRNA reductase.
GN VW10254.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.S.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016797; AAC08790.1; -.
KW Complete proteome.
SQ SEQUENCE 419 AA; 46057 MW; 834113C720656CBD CRC64;

Query Match 24.8%; Score 646; DB 16; Length 419;
Best Local Similarity 36.6%; Pred. No. 5e-30;
Matches 154; Conservative 81; Mismatches 144; Indels 42; Gaps 9;

QY 81 TIAVIGLSVHTAPVEMREKLAVALAEALPRAIOELTSLNHIIEAAVLSTCNRMETIVVALS 140
DB 2 SLLAIGINHHTASVDLREKVAFGDPDKLANALQQLSQHEAVNGSVILSTCNRTVYCDVKS 61
QY 141 WNRGIREVVDWMSKSGIPASELREHLFILRSSDATHLFEVSAGLSLVLGEGOILAQV 200
DB 62 GARS--KVIDWLSOFHQIGLEELAPSLYVVEEQAAIRHLMRVSCGLDSLVLGEPQLGQV 119
QY 201 KQVRSQNGSGGLGNIDRMFKDAITAGKRVSETNISGAVSVSSAAVELA---LMLKP 257
DB 120 KQAVSDREQAQVDALEKLFQKTFVSAKRVRTETDIGNAVSVAYAACTLAKHIFESLE 179
QY 258 KSEALSARMLLIGAGKMGKLVKHLVAGCKVWVNVRSVERVDAIREEMKDIEIVRPL 317
DB 180 KS-----TVLLVAGAGETIELVAKHANGCKSMIVANTRERALTALAQF-DAQVI--SL 231
QY 318 SDMYQAAAEADVFTSTAS-----ETSLFAKHAEALPPVSDTMGGVRLFVDISV 367
DB 232 QEIPNHAKADIVISSTASPLPIGKMGVETALKQRHQP-----MLLVDDIAY 279
QY 368 PRNVSACVSEVGARYVVDLKEVVEANKEDRLRKAMEAQIITEELRRFEAWRDSLET 427
DB 280 PRDVEAQVGLNDAYLTVDDLSQIIDSNIQORVQVEAIQAEIQAESAAFWMLRSLOA 339
QY 428 VPTIKKLRSYADIRIASELEKLCQKVGEDALTKMRRATEELSTGIVNKLKPLQLH 485
DB 340 VDSIREYQSANETREDLLSKALLSLRSGSDP-----EKVLRLELSNRLTNKLIHAPTRAL 394
QY 486 R 486
DB 395 Q 395

RESULT 10
Q8GCB0
ID Q8GCB0 PRELIMINARY; PRT; 446 AA.
AC Q8GCB0;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Glutamyl-tRNA reductase.
GN HWA.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM509;
RA Raux E., Leech H.K., Beck R., Schubert H.L., Santander P.J.,

```

RA Roesner C.A., Scott A.I., Thernes C., Rambach A., Martens J.H.,  
 RA Barg H., Jahn D., Warren M.J.;  
 RT "Identification and functional analysis of enzymes required for  
 RT precorrin-2 dehydrogenation and metal ion insertion in the  
 RT biosynthesis of sirohaem and cobalamin in *Bacillus megaterium*.";  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ508220; CAD48144.1;  
 SQ SEQUENCE 446 AA; 49838 MW; 8A86F67822352A02 CRC64;

Query Match 23.7%; Score 618.5; DB 2; Length 446;  
 Best Local Similarity 33.6%; Pred. No. 2.2e-28;  
 Matches 148; Conservative 101; Mismatches 177; Indels 15; Gaps 7;

Qy 82 IAVIGLSVHTAPVEMREKLAVALPRAIQELTSLNHIEAAVLSTCNRMIEIYVALSW 141  
 Db 3 IIAVGLNFTAPVETREKLSNFQELASAMKTLSCQKSIENIIVSTCNRTIEIYAVVDOL 62  
 Qy 142 NRGIREVVDWMSKSGIPASELRHFLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQVK 201  
 Db 63 HTGRYYVKAFLAEWFGIDKEEFPYLTIVENDGAIEHLRYVACGLDSMVIGETQILQVQR 122  
 Qy 202 QVVRSGQSGGLGKNIIDRMFKDAITAGKRVRSSTNISSGAVSVSSAAVELALMKLPKSEA 261  
 Db 123 SSFLAQAEETIGTVFNOLFQAVTLAKKAHETEIGANAVSVSAAVELA--KKIFGDL 180  
 Qy 262 LSARMLLIGAGKMGKLVKHLVAKGCKVWVNRVSVERVDAIREEMKDIEIVYRPLSDMY 321  
 Db 181 SSKVLLIGAGKMGQLAVQNLVSGAKKVVVNRITFEKAQLANRFSG---EAKPFDQLQ 237  
 Qy 322 QAAABADVFTSTASSETSLFAKEHAELPPVSDTMGGVRLF--VDISVPRNVSAVCSEVGA 380  
 Db 238 HALSEADILISSTGANDYVTVK---QMSAEARTKGRPLFVVDIAVRDLDPDELDEL 294  
 Qy 381 ARVYVVDLKEVVEANKEDRLKAMEAQTITTELRFRFEAWRDSLETVPITKILRSYADR 440  
 Db 295 VFLYDIDLNGVNESLOKQAADEIEIMLEAEIVAFKSMGLTGLGVVPIVIALRQKALT 354  
 Qy 441 IRASELEKCLKVGEDALTKMRRRAIEELSTGIVNKLHGLPQLHRCDSRSLTDELTE 500  
 Db 355 IQAETMKSIDRLPP--LSERKRVNKHITSIIINQLLRDIPHPAK--ELAGEKHAEESE 411  
 Qy 501 NNHALNRMFSLDMKAEIIEQK 521  
 Db 412 ---LPMKIFNIEQVALQKEK 429

RESULT 11  
 QBEAR2  
 ID Q8EAR2 PRELIMINARY; PRT; 416 AA.  
 AC Q8EAR2;  
 DT 01-MAR-2003 (TremBrel. 23, Created)  
 DT 01-MAR-2003 (TremBrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TremBrel. 23, Last annotation update)  
 DE Glutamy1-CRNA reductase.  
 GN HEMA OR SO3834.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=2297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vanathavan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealeon K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 Shewanella oneidensis.";

RL Nat. Biotechnol. 20:1118-1123 (2002).  
 DR EMBL: AB015816; AAN56811.1;  
 DR TIGR: SO3834;  
 KW Complete proteome.  
 SQ SEQUENCE 416 AA; 45686 MW; 336289F059E50C57 CRC64;

Query Match 22.9%; Score 597; DB 16; Length 416;  
 Best Local Similarity 35.0%; Pred. No. 3.7e-27;  
 Matches 142; Conservative 86; Mismatches 162; Indels 16; Gaps 6;

Qy 81 TIAVIGLSVHTAPVEMREKLAVALPRAIQELTSLNHIEAAVLSTCNRMIEIYVALSW 140  
 Db 2 SLVAIGINHKTATVDLREKVAFPDKTHDAMKSLASRTSGEAVIVSTCNRTIELYC----- 57  
 Qy 141 WNRGIR-EVVDWMSKSGIPASELRHFLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQ 199  
 Db 58 -NNGDETDIIEMLBEYHGLEHODVAPCLYNYHQGEAVRHLMRVASGLDSLILGEPOILQ 116  
 Qy 200 QVVRSGQSGGLGKNIIDRMFKDAITAGKRVRSSTNISSGAVSVSSAAVELALMKLPKS 259  
 Db 117 VKQAFVKAEGTVALTIDRLFQNTFSVAKKVRTETETEGAAAASVFAAASMA--KHIFS 174  
 Qy 260 EALSARMLLIGAGKMGKLVKHLVAKGCKVWVNRVSVERVDAIREEMKDIEIVYRPLSD 319  
 Db 175 SLSTTKVLLIGAGETIELVAKHLKONGVASVNVANRTILERAQSMCEEFNATAITLAQIPD 234  
 Qy 320 MYQAAABADVFTSTASSETSLFAKEHAELPPVSDTMGGVRLFVDISVPRNVSAVCSEVG 379  
 Db 235 F--LPRADIVISSTASPLPILGKMGVE--KALKQRRHQPMLLVDIAVPRDIEPEAVDL 289  
 Qy 380 AARVYVVDLKEVVEANKEDRLKAMEAQTITTELRFRFEAWRDSLETVPITKILRSYADR 439  
 Db 290 DAFLYTVDDLSHSIIIEQNKASKEAAEQABLITBESQSHLFMEWVRSLSDSIREYRSQM 349  
 Qy 440 RIRASELEKCLKVGEDALTKMRRRAIEELSTGIVNKLHGLPQLH 485  
 Db 350 AIKDELVERALKLAQSGDTEQV---LVELANRLTNRLIHAPTAOL 392

RESULT 12  
 Q8EPM7  
 ID Q8EPM7 PRELIMINARY; PRT; 457 AA.  
 AC Q8EPM7;  
 DT 01-MAR-2003 (TremBrel. 23, Created)  
 DT 01-MAR-2003 (TremBrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TremBrel. 23, Last annotation update)  
 DE Glutamy1-CRNA reductase (EC 1.2.1.1-).  
 GN HEMA OR OB2070.  
 OS Oceanobacillus iheyensis.  
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.  
 OX NCBI\_TaxID=182710;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HTE831 / DSM 14371 / JCM 11309;  
 RX MEDLINE=22220767; PubMed=12235376;  
 RA Takami H., Takaki Y., Uchiyama I.;  
 RT "Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.";  
 RL Nucleic Acids Res. 30:3927-3935 (2002).  
 DR EMBL: AP004599; BAC14026.1;  
 KW Oxidoreductase; Complete proteome.  
 SQ SEQUENCE 457 AA; 51544 MW; 2D936FCD7D9DFF4A CRC64;

Query Match 22.6%; Score 588; DB 16; Length 457;  
 Best Local Similarity 32.7%; Pred. No. 1.4e-26;  
 Matches 147; Conservative 90; Mismatches 195; Indels 18; Gaps 8;

Qy 82 IAVIGLSVHTAPVEMREKLAVALPRAIQELTSLNHIEAAVLSTCNRMIEIYVALSW 141  
 Db 3 ILKGVFNKTPVDIREKFTFSDSLQDAMVELKNQKSLIEDVIISTCNRTIEIYAVVDQL 62  
 Qy 142 NRGIREVVDWMSKSGIPASELRHFLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQVK 201

```
Db 63 HTGRYVVKOPLSNWFGIEKEEFSTYLRITEDDCAMEHLPRVSTGLDSVWLGTEQILQGVK 122
QY 202 QVVRSGNSGGLKNIDRMFKDAITAGKVRSETNISSGAVSVSSAAVELALMKLPKSEA 261
Db 123 QAFNLSQOVNTTGTIFNEUFPKAITFGKAHKEAJTAJEHAHSVSYAAVELA--KKIFGDL 180
QY 262 LSARMLLIGAGKMGKLVKHLVAKGCKVNVNRSVERVDAIREEMK-DIEIVYRPLSDM 320
Db 181 QEKHVAILGAGKMGKLAANKIQSGGATKITVNVNRTLENANEMAEKNADVE-----SIDQL 236
QY 321 YQAAAEADVFTSTASSETSLFAKEHAEALPPVSDTGGVRLFVDISVPRNVSAACVSEVGA 380
Db 237 PVILQADILISSTGADSVVTKEMMEKVQ--KORKGRALFLVDIAVPRDMDPAISELEN 294
QY 381 ARVYVDDILKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETVTIKKLSYADR 440
Db 295 VFYLDIDNLOHIVDDNLEGRKQAKIELLIEEIVTFKEWLKTLGVIPVISALROKALT 354
QY 441 IRASELEKCKLOKVGEDALTKMKRRAIEELSTGIVNKLHGPLQHLRCDGSDSRTLDITLE 500
Db 355 IQAETWQSIERKIPN--LTDREKVLNKHTKSIIINQLLEPVTQAK----EPAGKDNARD 408
QY 501 NMHALNMFSLDMKEAIIQKIK-AKVEKT 529
Db 409 SLQLPFINIFGI--EEBVKEELVKHAKNET 436

RESULT 13
Q8FI03
ID Q8FI03 PRELIMINARY; PRT; 484 AA.
AC Q8FI03;
DT 01-MAR-2003 (T-REMBLrel. 23, Created)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (T-REMBLrel. 23, Last annotation update)
DE Glutamy1-trRNA reductase (EC 1.2.1.1.).
GN HEMA OR C1668.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J.G., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RL of uropathogenic Escherichia coli."
DR Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AF016759; AAN80133.1; -.
KW Oxidoreductase, Complete proteome.
SQ SEQUENCE 484 AA; 53679 MW; 462CAB49487DC59 CRC64;

Query Match 21.9%; Score 570.5; DB 16; Length 484;
Best Local Similarity 33.7%; Pred. No. 1.6e-25;
Matches 141; Conservative 88; Mismatches 171; Indels 19; Gaps 8;

QY 81 TIATVGLSVHTAPVEMREKLAVALPRAIQELTSLNHIIEAAVLSTCNRMIEIYVALS 140
Db 68 TLLALGINHTKAPVSLRERSVSPDKLDQALDSLLAPQVQGVVLTSCNRTLYLSVEE 127
QY 141 WNRGIREVVDMWKKSGIPASERLHFLTRSSDATHLFEVSAGLDSVLGEGQILAQV 200
Db 128 RDLQEARLRLUCDHYHNLNEDDLRSLYWHQNDVASHLMRVASGLDSVLGEPQILGV 187
QY 201 KOVVRSGNSGGLKNIDRMFKDAITAGKVRSETNISSGAVSVSSAAVELALMKLPKSEA 260
Db 188 KKAFADSKQGHMKASELERMFQKSFSAKRVATETDIDGASVSVFAACTLARQIF---E 244
QY 261 ALS-ARMILLIGAGKMGKLVKHLVAKGCKVNVNRSVERVDAIREEMKDIEIVYRPLSD 319
```

```
Db 245 SLSTVTVLVGAGETIELVARHLREHKVQKMIANRTTREAQIADDEV-GAEVI--ALSD 301
QY 320 MYQAAAEADVFTSTASSETSLFAKEHAEALPPVSDTGGVRLFVDISVPRNVSAACVSEVG 379
Db 302 IDERLEADIIISSTASPLPIIGKGVVER--ALKSRNQPMILLVIAVPRDVEPEVGKLA 359
QY 380 AARVYVDDILKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETVTIKKLSYAD 439
Db 360 NAYLSVDDLOQSIISHLAQKAAVAEETIIVAQEASEFMAMURAOSASSETIRDSQAE 419
QY 440 RIRASELEKCKLOKV--GEDALTKMKRRAIEELSTGIVNKLHGPL--LQHLRCDGSDSR 493
Db 420 QYRDELTAALAAALSGGDAQT-----IMQDLAKLTLNRLIHAPTLSLQQAARDGNER 473

RESULT 14
Q8CNV6
ID Q8CNV6 PRELIMINARY; PRT; 448 AA.
AC Q8CNV6;
DT 01-MAR-2003 (T-REMBLrel. 23, Created)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (T-REMBLrel. 23, Last annotation update)
DE Glutamy1-trRNA reductase.
GN SE1347.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016748; AAO04946.1; -.
KW Complete proteome.
SQ SEQUENCE 448 AA; 50151 MW; F89F44784F3256A2 CRC64;

Query Match 20.9%; Score 543.5; DB 16; Length 448;
Best Local Similarity 30.5%; Pred. No. 5.6e-24;
Matches 137; Conservative 103; Mismatches 182; Indels 27; Gaps 9;

QY 85 IGLSVHTAPVEMREKLAVALPRAIQELTSLNHIIEAAVLSTCNRMIEIYVALSWNRG 144
Db 6 ISINHTADVLTUREQVAFRDDALRLAHEDLYETKAILLENVILSTCNRTTEVYAIVDQVHTG 65
QY 145 IREVVDWKKSGIPASERLHFLTRSSDATHLFEVSAGLDSVLGEGQILAQVQV 204
Db 66 RYIQRFLARSFGFEVDIDKMSQVKVGDDAVEHLLRVTSGLDSIVLGETQILQMRDAP 125
QY 205 RSGQNSGGLKNIDRMFKDAITAGKVRSETNISSGAVSVSSAAVELALMKLPKSEALSA 264
Db 126 FLAQNLTGTTGTIFNHLFKQAITFAKKAHSETDIADNAVSVSYAAVELAKKVFGLK--SK 183
QY 265 RMLLIGAGKMGKLVKHLVAKGCKVNVNRSVERVDAIREEMKDIEIVYRPLSDMYQAA 324
Db 184 HAVVLIGAGEMELSILNLLGSGISNVTIVNRTLSKAKILAEKH---NVSYDSLALPSLL 240
QY 325 AEADVFTSTASSETSLFAKEHAEALPPVS--DTMGVRLFVDISVPRNVSAACVSEVGAAR 382
Db 241 ETTDVISSTSAEDYIITNSVMVKTISERKLDLSL---VLIDIAVPRDIEPIDAITNIF 296
QY 383 VYVDDILKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETVTIKKLSYAD 439
Db 297 NYDVDDLDKLDVANDLRER---QLAAETTAGIPEIDSHNENWVMVGLVVPVIRALREKAM 353
QY 440 RIRASELEKCKLOKVGEDALTKMKRRAIEELSTGIVNKLHGPLQHLRCDGSDSRTLDTEL 499
Db 354 NIQETMESIDRKLPD--LSERERKVISKHTKSIINQMLKPIQAKELSTDDK----SN 407
QY 500 ENMHALNMFSLDMKEAIIQKIKAKVEK 528
Db 408 EKLELFQWIFDIEAE---DPREKAKLEK 432
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 12:53:42 ; Search time 47 Seconds  
(without alignments)  
1793.273 Million cell updates/sec

Title: US-10-018-902-4

Perfect score: 2606

Sequence: 1 MATTTTATTAATAAATAATKAP.....DMEKALIEQIKAKVETQN 531

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2606	100.0	531	22 AAB31960	Amino acid sequenc
2	2272	87.2	480	22 AAB31962	Amino acid sequenc
3	2091	80.2	454	22 AAB31968	Amino acid sequenc
4	1783.5	68.4	536	22 AAB31964	Amino acid sequenc
5	1770.5	67.9	543	23 ABB31383	Herbicidally activ
6	1677	64.4	530	23 ABB30929	Herbicidally activ
7	1540	59.1	467	22 AAB31966	Amino acid sequenc
8	1334	51.2	524	23 ABB31959	Herbicidally activ
9	637.5	24.5	415	21 AAY75260	Neisseria meningit

10	632.5	24.3	415	21	AAY75259	Neisseria gonorrhoe
11	632.5	24.3	415	24	ABP78975	N. gonorrhoeae ami
12	628.5	24.1	415	21	AAY75261	Neisseria meningit
13	594.5	22.8	426	16	AAW74619	Xanthomonas L-glut
14	543.5	20.9	448	22	AAG82317	S. epidermidis ope
15	543.5	20.9	448	23	ABP38210	Staphylococcus epi
16	506	19.4	435	23	ABBA4846	Listeria monocytog
17	483.5	18.6	463	22	AAG90216	C glutamicum prote
18	428.5	16.4	449	19	AAW98821	H. pylori GHP0 142
19	388	14.9	87	22	AAB31961	Amino acid sequenc
20	282	10.8	63	22	AAB31959	Amino acid sequenc
21	219.5	8.4	384	22	AAU67473	Propionibacterium
22	206.5	7.9	341	20	AAY35341	Chlamydia pneumonia
23	195.5	7.5	348	20	AAY37517	Protein involved i
24	145.5	5.6	1764	10	AAW91672	Primary amino acid
25	139.5	5.4	975	19	AAW72746	Drosophila kinesin
26	139.5	5.4	975	22	ABB63485	Drosophila melanog
27	139	5.3	1275	22	AAW78921	Human protein SEQ
28	139	5.3	1301	22	AAW79905	Human protein SEQ
29	139	5.3	2017	22	ABG66301	Novel human diagno
30	138	5.3	1725	21	AAW19800	Mouse laminin 2 ma
31	138	5.3	1725	21	AAW48451	Mouse laminin 8 po
32	138	5.3	1725	23	ABB81593	Mouse laminin 10 s
33	138	5.3	1786	21	AAW19799	Mouse laminin 2 be
34	138	5.3	1786	21	AAW48450	Mouse laminin 8 po
35	138	5.3	1786	23	ABB81592	Mouse laminin 10 s
36	136.5	5.2	1776	19	AAW50894	Mouse laminin B1 c
37	135	5.2	3338	22	ABG09919	Novel human diagno
38	133	5.1	2563	23	AAU10873	Penicillium citrin
39	132.5	5.1	871	23	ABP39134	Staphylococcus epi
40	132	5.1	727	21	AAW47308	Arabidopsis thalia
41	129.5	5.0	1123	22	ABG12514	Novel human diagno
42	129.5	5.0	1123	22	ABG14529	Novel human diagno
43	129	5.0	648	23	ABP73703	Candida albicans e
44	129	5.0	1177	22	AAW96721	Putative P. abyssi
45	128.5	4.9	2757	22	ABG09921	Novel human diagno

#### ALIGNMENTS

#### RESULT 1

AAB31960  
ID AAB31960 standard; Protein; 531 AA.

AC AAB31960;  
AC

DT 15-MAY-2001 (first entry)  
DT

DE Amino acid sequence of a corn Glu-tRNA reductase enzyme.  
DE

KW Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;  
KW tetrapyrrolic pigment; chlorophyll; transgenic plant.  
KW

OS Zea mays.  
OS

PN W0200109304-A2.  
PN

XX 08-FEB-2001.  
XX

PF 28-JUL-2000; 2000WO-US21008.  
PF

PR 30-JUL-1999; 99US-0146600.  
PR

XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX

PI Cahoon RE, Gutteridge S, Harvell LT, Rafalski JA, Tao Y, Weng Z;  
PI

XX WPI; 2001-159865/16.  
XX

DR N-PSDB; AAF54814.  
DR

XX New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme  
XX useful for producing large amounts of the encoded polypeptides used in  
PT

PT screening compounds for potential herbicidal activity -  
 PS Disclosure; Page 48-49; 77pp; English.

CC The present sequence represents a Glu-tRNA reductase, which is an  
 CC aminolevulinic acid biosynthetic enzyme. The enzyme converts  
 CC Glu-tRNA-Glu to glutamate 1-semialdehyde (GSA) with the concomitant  
 CC release of tRNA-Glu. GSA aminotransferase then converts GSA to  
 CC 5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of  
 CC tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding  
 CC aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs  
 CC and genes encoding homologous proteins from the same or other plant  
 CC species, for creating transgenic plants in which the polypeptides are  
 CC present at higher or lower levels than normal or in cell types or  
 CC developmental stages in which they are not normally found, for  
 CC overexpression in bacterial or yeast hosts to efficiently produce  
 CC large amounts of the encoded polypeptides which could then be used for  
 CC screening different compounds for potential herbicidal activity, and  
 CC as hybridisation probes and amplification primers.

XX Sequence 531 AA;

Query Match 100.0%; Score 2606; DB 22; Length 531;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-213;  
 Matches 531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATTTATTAATAATTAAPRGSSSALCQVAGGRRRRSGVVRCDAAAGVEAQAQAA 60  
 Db |||||  
 Qy 1 MATTTATTAATAATTAAPRGSSSALCQVAGGRRRRSGVVRCDAAAGVEAQAQAA 60  
 Db |||||  
 Qy 61 SVALEQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALPRAIQELTSLNHI 120  
 Db |||||  
 Qy 61 SVALEQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALPRAIQELTSLNHI 120  
 Db |||||  
 Qy 121 EEAVALSTCNRMETVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDATHLP 180  
 Db |||||  
 Qy 121 EEAVALSTCNRMETVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDATHLP 180  
 Db |||||  
 Qy 181 EVSAGLSLVLEGQILAQVQVVRSGQNSGLGNIDRMFKDAITAGKRVSETNISSG 240  
 Db |||||  
 Qy 181 EVSAGLSLVLEGQILAQVQVVRSGQNSGLGNIDRMFKDAITAGKRVSETNISSG 240  
 Db |||||  
 Qy 241 AVSVSSAAVELALMKLPKSEALSARMLLIGAKGKGLVIRKHLVAKGCKVVRNRSERV 300  
 Db |||||  
 Qy 241 AVSVSSAAVELALMKLPKSEALSARMLLIGAKGKGLVIRKHLVAKGCKVVRNRSERV 300  
 Db |||||  
 Qy 301 DAIREEMKDIEIVYRPLSDMYQAAAEADVFTSTASSETSLFAKEHAELPPVSDTMGGVR 360  
 Db |||||  
 Qy 301 DAIREEMKDIEIVYRPLSDMYQAAAEADVFTSTASSETSLFAKEHAELPPVSDTMGGVR 360  
 Db |||||  
 Qy 361 LFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTITTEELRFEA 420  
 Db |||||  
 Qy 361 LFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTITTEELRFEA 420  
 Db |||||  
 Qy 421 WRDSLETVPITIKLRSYADRIASELEKCLKQVGEDALTQKORRAIEELSTGIYVKNLHG 480  
 Db |||||  
 Qy 421 WRDSLETVPITIKLRSYADRIASELEKCLKQVGEDALTQKORRAIEELSTGIYVKNLHG 480  
 Db |||||  
 Qy 481 PLOHLRCDGSDSRDLTLENHMLNRMFSLDMEKAIIEQIKAKVEKTON 531  
 Db |||||  
 Qy 481 PLOHLRCDGSDSRDLTLENHMLNRMFSLDMEKAIIEQIKAKVEKTON 531  
 Db |||||

RESULT 2

AAB31962  
 ID AAB31962 standard; Protein; 480 AA.

XX  
 AC AAB31962;

XX 15-MAY-2001 (first entry)

XX Amino acid sequence of a rice Glu-tRNA reductase enzyme.

KW Glu-tRNA reductase; aminolevulinic acid; 5-aminolevulinic acid; ALA;  
 KW tetrapyrrolic pigment; chlorophyll; transgenic plant.  
 XX Oryza sativa.  
 XX WO200109304-A2.  
 XX 08-FEB-2001.  
 XX 28-JUL-2000; 2000WO-US21008.  
 XX 30-JUL-1999; 99US-0146600.  
 XX (DUPO) DU PONT DE NEMOURS & CO E I.  
 XX Cahoon RE, Gutteridge S, Harvell LT, Rafaleki JA, Tao Y, Weng Z;  
 PI MPI; 2001-159865/16.  
 DR N-PSDB; AAF54816.  
 XX

PT New nucleic acids encoding an aminolevulinic acid biosynthetic enzyme  
 PT useful for producing large amounts of the encoded polypeptides used in  
 PT screening compounds for potential herbicidal activity -  
 XX Disclosure; Page 52-54; 77pp; English.

CC The present sequence represents a Glu-tRNA reductase, which is an  
 CC aminolevulinic acid biosynthetic enzyme. The enzyme converts  
 CC Glu-tRNA-Glu to glutamate 1-semialdehyde (GSA) with the concomitant  
 CC release of tRNA-Glu. GSA aminotransferase then converts GSA to  
 CC 5-aminolevulinic acid (ALA). ALA is used in the biosynthesis of  
 CC tetrapyrrolic pigments such as chlorophyll. Nucleic acids encoding  
 CC aminolevulinic acid biosynthetic enzymes are useful for isolating cDNAs  
 CC and genes encoding homologous proteins from the same or other plant  
 CC species, for creating transgenic plants in which the polypeptides are  
 CC present at higher or lower levels than normal or in cell types or  
 CC developmental stages in which they are not normally found, for  
 CC overexpression in bacterial or yeast hosts to efficiently produce  
 CC large amounts of the encoded polypeptides which could then be used for  
 CC screening different compounds for potential herbicidal activity, and  
 CC as hybridisation probes and amplification primers.

XX Sequence 480 AA;

Query Match 87.2%; Score 2272; DB 22; Length 480;  
 Best Local Similarity 95.2%; Pred. No. 1.3e-184;  
 Matches 457; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

Qy 52 QAQAVAKAAVAALAEQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALPRAI 111  
 Db |||||  
 Qy 1 QAQAVAKAAVAALAEQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALPRAI 60  
 Db |||||  
 Qy 112 QELTSLNHIIEAAVLSTCNRMETVVALSWNRGIREVVDWMSKSGIPASELREHLFILR 171  
 Db |||||  
 Qy 61 SELTSLNHIIEAAVLSTCNRMETVVALSWNRGIREVVDWMSKSGIPASELREHLPMLR 120  
 Db |||||  
 Qy 172 SSDATHRLFEVSSAGLSLVLEGQILAQVQVVRSGQNSGLGNIDRMFKDAITAGKRV 231  
 Db |||||  
 Qy 121 DSDATHRLFEVSSAGLSLVLEGQILAQVQVVRSGQNSGLGNIDRMFKDAITAGKRV 180  
 Db |||||  
 Qy 232 RSETNISSGAVSVSSAAVELALMKLPKSEALSARMLLIGAKGKGLVIRKHLVAKGCKVV 291  
 Db |||||  
 Qy 181 RSETNISSGAVSVSSAAVELALMKLPKSEALSARMLLIGAKGKGLVIRKHLVAKGCKVV 240  
 Db |||||  
 Qy 292 VVNRSEVRVDAREEMKDIEIVYRPLSDMYQAAAEADVFTSTASSETSLFAKEHAELPP 351  
 Db |||||  
 Qy 241 VVNRSEVRVDAREEMKDIEIVYRPLSDMYQAAAEADVFTSTASSETSLFAKEHAELPP 300  
 Db |||||  
 Qy 352 VSDTMGGVRLFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTII 411  
 Db |||||  
 Qy 301 ISDAMGGVRLFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTII 360  
 Db |||||  
 Qy 412 TEELRFEAWRDSLETVPITIKLRSYADRIASELEKCLKQVGEDALTQKORRAIEELST 471  
 Db |||||









PI Tettelin H, Venter JC;  
 XX WPI: 2000-062150/05.  
 DR N-PSDB; AA254022.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics -  
 XX  
 PS Claim 2; Page 996; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254615 to AA254615  
 CC represent novel Neisseria meningitis and N. gonorrhoea polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254615 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 415 AA;  
 Query Match 24.5%; Score 637.5; DB 21; Length 415;  
 Best Local Similarity 34.3%; Pred. No. 1.5e-45;  
 Matches 149; Conservative 96; Mismatches 163; Indels 27; Gaps 9;  
 QY 82 IAVIGLSVHTAPVEMREKLAVALAEELPRAIQELTSLNHIIEAAVLSTCNRMIEIYVVALSW 141  
 Db 3 LTAVGLNHQTAFLSIREKLAFAAAALPRAVNLARSNAATEAVILSTCNRTLYCVGDS- 61  
 QY 142 NRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSLVLGEQILAQVK 201  
 Db 62 ----EIIIRWLADYHSLPIEIRPYLYALDMQETVRHAFRVACGLDSVLGEPQILGQIK 117  
 QY 202 QVVRSGQSGGLGNIDRMFKDAITAGKRVSRSETNISSGAVSVSSAAVELALMKLPKSEA 261  
 Db 118 DAVRAAQEQESMGKLNALFQKTSVAKEVRTDTAVGENSVSMASVVKLAEQIFPDIGD 177  
 QY 262 LSARMLLIGAKMGKLVKHLVAKGCKKVVVNRVSVERVDAIREMKD-IBIVYRP--LS 318  
 Db 178 LN--VLFIGAGEMIELVATYFAAKNPRMTVANRTLARA----QELCDKLGVAEPCLLS 231  
 QY 319 DMYQAAAEADVFTSTASETSLFAKEHAELPPVSDTMGGVRLP-VDISVPRNVACVSE 377  
 Db 232 DLPAILHDYDVVSVSTASQLPVIGKGMVERALKQKQSM---PLFMLDLAVPRDIEAEVGD 288  
 QY 378 VGAARVYVNDLKEVVEANKEDRLRKAMEAQTIIITEELRRFEAWRDSLETPTIKLRSY 437  
 Db 289 LNDAYLYTVDDMNVIVSGKEARQAKAAATAATLVSEKVAEFVROOQGRQSVPLIKALRDE 348  
 QY 438 ADRTASELEKCLKQVGEDALTQKORRAIBELSTGIVNKLHGLPQHLRCDSGDSRTLDE 497  
 Db 349 GEKARKQVLENAMQLAKGATAEEV---LERLSVQLTNKLHLSPTQTLINKAGEEDKDL-- 403  
 QY 498 TLENWHLNRPFSLD 512  
 Db 404 ----VHAQAQIYHLD 414  
 RESULT 10  
 AA2575259  
 ID AA2575259 standard; Protein; 415 AA.  
 XX  
 AC AA2575259;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria gonorrhoea ORF 622 protein sequence SEQ ID NO:1992.  
 XX

Neisseria meningitidis; Neisseria gonorrhoea; antigen; vaccine;  
 antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 antibacterial; gene therapy.  
 Neisseria gonorrhoea.  
 WO9957280-A2.  
 11-NOV-1999.  
 30-APR-1999; 99WO-US09345.  
 01-MAY-1998; 98US-0083758.  
 31-JUL-1998; 98US-0094869.  
 02-SEP-1998; 98US-0098994.  
 02-SEP-1998; 98US-0099062.  
 09-OCT-1998; 98US-0103749.  
 09-OCT-1998; 98US-0103794.  
 09-OCT-1998; 98US-0103796.  
 25-FEB-1999; 99US-0121528.  
 (CHIR) CHIRON CORP.  
 (GENO-) INST GENOMIC RES.  
 Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 Petersen J, Pizzi M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 Tettelin H, Venter JC;  
 WPI: 2000-062150/05.  
 N-PSDB; AA254021.  
 Novel Neisserial polypeptides predicted to be useful antigens for  
 vaccines and diagnostics -  
 Claim 2; Page 995; 1453pp; English.  
 AA253015 to AA254536, AA254577 to AA254615, and AA254615 to AA254615  
 represent novel Neisseria meningitis and N. gonorrhoea polynucleotides  
 and polypeptides. AA254537 to AA254576 and AA254616 to AA254615 represent  
 PCR primers used in the exemplification of the present invention. The  
 polypeptides, the polynucleotides, antibodies and compositions of  
 the invention can be used as vaccines, as diagnostic reagents, and as  
 immunogenic compositions. The polypeptides can be used in the  
 manufacture of medicaments for treating or preventing infection due to  
 Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 presence of Neisseria bacteria, or to raise antibodies. They may also  
 be used to screen for agonists or antagonists, which may themselves  
 have use as antibacterial agents. The polynucleotides of the invention  
 may also be used in gene therapy protocols.

Query Match 24.3%; Score 632.5; DB 21; Length 415;  
 Best Local Similarity 34.0%; Pred. No. 3.9e-45;  
 Matches 148; Conservative 96; Mismatches 164; Indels 27; Gaps 9;  
 QY 82 IAVIGLSVHTAPVEMREKLAVALAEELPRAIQELTSLNHIIEAAVLSTCNRMIEIYVVALSW 141  
 Db 3 LTAVGLNHQTAFLSIREKLAFAAAALPRAVNLARSNAATEAVILSTCNRTLYCVGDS- 61  
 QY 142 NRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSLVLGEQILAQVK 201  
 Db 62 ----EIIIRWLADYHSLPIEIRPYLYALDMQETVRHAFRVACGLDSVLGEPQILGQIK 117  
 QY 202 QVVRSGQSGGLGNIDRMFKDAITAGKRVSRSETNISSGAVSVSSAAVELALMKLPKSEA 261  
 Db 118 DAVRAAQEQESMGKLNALFQKTSVAKEVRTDTAVGENSVSMASVVKLAEQIFPDIGD 177  
 QY 262 LSARMLLIGAKMGKLVKHLVAKGCKKVVVNRVSVERVDAIREMKD-IBIVYRP--LS 318  
 Db 178 LN--VLFIGAGEMIELVATYFAAKNPRMTVANRTLARA----QELCDKLGVAEPCLLS 231  
 QY 319 DMYQAAAEADVFTSTASETSLFAKEHAELPPVSDTMGGVRLP-VDISVPRNVACVSE 377



Db 232 DLPAILHDYDVVVSTASQPIVKGKMERALKQRQSM---PLFMLDLAVPRDIEAEVGD 288  
 Qy 378 VGAARVYVVDLKEYVEANKEDRLRKAMEAQTIIITEELRRFEAWRDSLETPTIKLRSY 437  
 Db 289 LNDAYLYTVDDMVNIVQSGKEARQKAAAAAETLVSEKVAEFVROOQGRQSVPLIKALRDE 348  
 Qy 438 ADRIASELEKCLKQVGEDALTCKMRRRAIBELSTGIYNKLLHGLPQHLRCDGSDSRITLDE 497  
 Db 349 GEKARKQVLENAMKQLAGATAEEV---LERLSVQLTNKLLHSPQTTLNKAGEEDKOL-- 403  
 Qy 498 TLENMHALNRMFSLD 512  
 Db 404 ----VHAVAQIYHLD 414

## RESULT 11

ABP78975  
 ID ABP78975 standard; Protein; 415 AA.

AC ABP78975;  
 XX 07-MAR-2003 (first entry)  
 DT N. gonorrhoeae amino acid sequence SEQ ID 4480.  
 DE Antibacterial; infection; vaccine; gene therapy.  
 KW Neisseria gonorrhoeae.

OS WO200279243-A2.  
 PN 10-OCT-2002.  
 PD 12-FEB-2002; 2002WO-IB02069.  
 XX 12-FEB-2001; 2001GB-0003424.

XX (CHIR-) CHIRON SPA.

XX Fontana MR, Pizza M, Massignani V, Monaci E;  
 XX WPI; 2003-058415/05.  
 DR N-PSDB; ABZ39945.  
 XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection -  
 PT Disclosure; Page 516; 815pp; English.

XX The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention.

XX Sequence 415 AA;

Query Match 24.3%; Score 632.5; DB 24; Length 415;  
 Best Local Similarity 34.0%; Pred. No. 3.9e-45;  
 Matches 148; Conservative 96; Mismatches 164; Indels 27; Gaps 9;

Qy 82 IAVIGLSVHTAPVEMREKLAVALAEELMPRAQELTSLNHIIEAAVLSTCNRMETVVALSW 141  
 Db 3 LTAVGLNHQTAPLSIREKLAFAAALPEAVRNLRASNAATEAVILSTCNRTLEYCVGDS- 61

Qy 142 NRGIREVVDNMKSGIPASELREHLFILRSSDATRHLFEVSAGLSLVLGEGQILAOVK 201  
 Db 62 ----BEIRLADYHSLFIEIRPLYLTDQETVRHAFRVACGLDSMWLGEFQILGQIK 117

Qy 202 QVTRSGNSGGLGNKIDRMFKDAITAGKVRSETNISSGAVSVSSAAVELALMKLPKSEA 261  
 Db 118 DAVRAAQEQESMGAKNALFOKTPSVAKEVRTDTAVGENSVSMASAVKLAEQIFPDIGD 177  
 Qy 262 LSABMLLIGACKMKLVKHLVAKGCKKVVVVRVSVERVDAIREMKD-IEIVYRP--LS 318  
 Db 178 LN--VLFIGAGEMELVATYFAAKNPRDLMTVANRTLARA----QELCDKUGVNAEPCLLS 231  
 Qy 319 DMYQAAAADVVFTSTASSETSLFAKEHAELPPVSDTMGGVRLF-VDISVPRNVSACVSE 377  
 Db 232 DLPAILHDYDVVVSTASQPIVKGKMERALKQRQSM---PLFMLDLAVPRDIEAEVGD 288  
 Qy 378 VGAARVYVVDLKEYVEANKEDRLRKAMEAQTIIITEELRRFEAWRDSLETPTIKLRSY 437  
 Db 289 LNDAYLYTVDDMVNIVQSGKEARQKAAAAAETLVSEKVAEFVROOQGRQSVPLIKALRDE 348  
 Qy 438 ADRIASELEKCLKQVGEDALTCKMRRRAIBELSTGIYNKLLHGLPQHLRCDGSDSRITLDE 497  
 Db 349 GEKARKQVLENAMKQLAGATAEEV---LERLSVQLTNKLLHSPQTTLNKAGEEDKOL-- 403  
 Qy 498 TLENMHALNRMFSLD 512  
 Db 404 ----VHAVAQIYHLD 414

## RESULT 12

AA75261  
 ID AA75261 standard; Protein; 415 AA.

AC AA75261;

XX 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 622 protein sequence SEQ ID NO:1996.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy.

XX Neisseria meningitidis.

XX WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;  
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX WPI; 2000-062150/05.

DR N-PSDB; AAZ54023.

XX Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 PT Claim 2; Page 997; 1453pp; English.

PS AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ74253 to AAZ75941

CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides

CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX  
 SQ Sequence 415 AA;

Query Match 24.1%; Score 628.5; DB 21; Length 415;  
 Best Local Similarity 34.0%; Pred. No. 8.6e-45;  
 Matches 148; Conservative 95; Mismatches 165; Indels 27; Gaps 9;

Qy 82 IAVIGLSVHTAPVEMREKLAVALAEELMPRAIOELTSLNHIEEAAVLSTCNRMIEYVVALSW 141  
 Db 3 LTAUGLNHQTAPLSIREKLAFAAACLPFAVRNLAARSNAATEAVILSTCNRTLCVGDG- 61  
 Qy 142 NRGTRVVDWMSKSGIPASELREHLFILRSSDATRHLPFVSAGLDSLVLGEGQILAQVK 201  
 Db 62 ----EEIIRWLADYHSLPIEIEISPYLYTLGMQETVRAFRVACGLDSHVLGEPQILGQIK 117  
 Qy 202 QVRSQNSGGGLGNIDRMFKDAITAGKRVSRSETNISSGAVSVSSAAVELALMKLPKSEA 261  
 Db 118 DAVRVAQESMGKKNALFQRTSVAKVEVRTDTAVGENSVSMASVVKLAEQIFPDIGD 177  
 Qy 262 LSARMLLIGAGKMGKLVIKHLVAKGCKKVVVNVRSVERVDAREEMKD-IEIVYRP--LS 318  
 Db 178 LN--VLFIGAGEMIELVATYPAKSPRLMTVANRTLABA---QELCDKLGVAEPCLLS 231  
 Qy 319 DMYQAAEADVVFTSASETSLFAKEHAELPPVSDTNGVGRFLP-VDISVPRNVSAVCSE 377  
 Db 232 DLPAILHEYDVVVSSTASQLPIVGKGMVERALKQKQSM--PLFMDLAVPRDIEAEVGD 288  
 Qy 378 VGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETPTTKLSRY 437  
 Db 289 LNDAYLYTVDDMNVIVSGKEARQKAAAEATLSEKVAEVRQOQGRQSVPLIRALRDE 348  
 Qy 438 ADRTRASELEKCLKVGEDALTKKWRRAIEELSTGIVNKLHGLHQLRCDGSDSRITLDE 497  
 Db 349 GEKARKQVLENAMQLAKGATAEEV---LERLSIQLTNKLHSTQTINKAGEEDKDL-- 403  
 Qy 498 TLENHMLNRNFSLD 512  
 Db 404 ----VHAVAQIYHLD 414

RESULT 13

AA82317  
 ID AAR74619 standard; Protein; 426 AA.

XX  
 AC AAR74619;

DT 02-NOV-1995 (first entry)

XX Xanthomonas L-glutamyl tRNA reductase.

XX Xanthomonas campestris; L-glutamyl tRNA reductase; vitamin B12;  
 KW glutamic acid-1-semialdehyde; 5-aminolevulinic acid (ALA); haem;  
 XX chlorophyll; phycobillin; recombinant production.

OS Xanthomonas campestris.

XX JP070311480-A.

XX 03-FEB-1995.

XX 27-JUL-1993; 93JP-0184709.

XX

PR 27-JUL-1993; 93JP-0184709.

XX (COSM-) COSMO SOGO KENKYUSHO KK.

XX WPI; 1995-109535/15.

XX N-PSDB; AAQ88151.

PT DNA fragment encoding L-glutamyl tRNA reductase - derived from  
 PT Xanthomonas sp.; useful for production of 5-amino:levulinic,  
 PT vitamin-B12, chlorophyll, etc.

XX Claim 2; Pages 7-8; 9pp; Japanese.

XX AAQ88151 encodes AAR74619 Xanthomonas campestris L-glutamyl tRNA  
 CC reductase. The DNA as part of an expression vector can be used to  
 CC transform a host cell, for the recombinant prodn. of the protein.  
 CC The protein can be used in the prodn. of vitamin B12, glutamic  
 CC acid-1-semialdehyde, 5-aminolevulinic, haem, chlorophyll and  
 CC phycobillin.

XX SQ Sequence 426 AA;

Query Match 22.8%; Score 594.5; DB 16; Length 426;  
 Best Local Similarity 34.4%; Pred. No. 7e-42;  
 Matches 150; Conservative 85; Mismatches 164; Indels 37; Gaps 11;

Qy 81 TIAVIGLSVHTAPVEMREKLAVALAEELMPRAIOELTSLNHIEEAAVLSTCNRMIEYVVALS 140  
 Db 2 TLVGLGNHQTAPVLDLERRAAFGDALPRALESRLALPQVSEALLSTCNRTLEYAMA-- 59  
 Qy 141 WNRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLPFVSAGLDSLVLGEGQILAQV 200  
 Db 60 --EEAHSVLTWLETH----APALSGYLYQHQAEEAVRHLPFRVATGLDSHVLGEPQILGQV 113  
 Qy 201 KOVRSQNSGGGLGNIDRMFKDAITAGKRVSRSETNISSGAVSVSSAAVELALMKLPKSE 260  
 Db 114 KDAAVARAHGTLGSLDRLFFOOTFSVAKKARTDTRVCANPVSVAATAVRLAQDSFARLN 173  
 Qy 261 ALSARMLLIGAGKMGKLVIKHLVAKGCKKVVVNVRSVERVDAREEMKDIEIVYRPLSDM 320  
 Db 174 --ESTVLLIGAGETIELAAKHLSEGRVRRLLIANRTLAHAQTLASQHGFPAL---PLTDL 228  
 Qy 321 YQAAAEADVFTSASETSLFAKEHAEL-AL-----PPVSDTNGVGRFLVDISVPRNVSA 374  
 Db 229 ERHIAEADVFTSASETSLFAKEHAEL-AL-----PPVSDTNGVGRFLVDISVPRNVSA 374  
 Qy 375 VSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETPTTKK 433  
 Db 281 VGELSDAYLYTVDDLERAVEDNRGRREAAQDAEALDLQVARYVETLQANARQAP-LKR 339  
 Qy 434 LRSYADRIRASELEKCLKVGEDALTKKWRRAIEELSTGIVNKLHGLHQLRCDGSDSR 493  
 Db 340 LRAFGDSTRDELLAKARQQLHNG---KPADEVLEQLAHALTNRLLHPTTAALR-----DA 391  
 Qy 494 TLDETLENHMLNRNMF 509  
 Db 392 ALANDLELTAAADRLF 407

RESULT 14

AA82317

ID AAG82317 standard; Protein; 448 AA.

XX  
 AC AAG82317;

DT 03-SEP-2001 (first entry)

XX S. epidermidis open reading frame protein sequence SEQ ID NO:1728.

DE Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 XX Staphylococcus epidermidis.

XX Staphylococcus epidermidis.





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 12:56:14 ; Search time 37 Seconds  
(without alignments)  
2669.117 Million cell updates/sec

Title: US-10-018-902-4  
Perfect score: 2606  
Sequence: 1 MATTTATTAATAATAATKAP.....DMEKALIEQKIKAVKXNTON 531

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	483.5	18.6	463	10	US-09-738-626-3970
2	428.5	16.4	449	12	US-09-882-227-508
3	427	16.4	569	15	US-10-156-761-12273
4	143	5.5	1258	15	US-10-156-761-10395
5	139.5	5.4	975	12	US-10-080-608A-19
6	139.5	5.4	975	12	US-10-080-608A-19
7	138	5.3	1725	15	US-10-037-685-108
8	138	5.3	1786	10	US-10-037-182-12
9	138	5.3	1786	15	US-09-938-275-7
10	137.5	5.3	1979	11	US-10-205-823-419
11	133	5.1	2563	11	US-09-836-705-46
12	130	5.0	975	12	US-10-094-749-2015
13	129	5.0	648	12	US-10-032-585-7540
14	128	4.9	869	9	US-09-815-242-5230
15	128	4.9	862	9	US-09-815-242-12526

16	128	4.9	999	15	US-10-156-761-14584	Sequence 14584, A
17	124.5	4.8	1178	15	US-10-128-714-8240	Sequence 8240, Ap
18	124.5	4.8	1566	15	US-10-136-761-14473	Sequence 14473, A
19	124	4.8	856	9	US-09-815-242-11134	Sequence 11134, A
20	124	4.8	945	9	US-09-745-763-191	Sequence 191, App
21	124	4.8	1000	12	US-10-012-697-1540	Sequence 1540, Ap
22	124	4.8	1711	10	US-09-771-161A-219	Sequence 219, App
23	124	4.8	1711	10	US-09-771-161A-220	Sequence 220, App
24	119	4.6	970	15	US-10-078-531-4	Sequence 4, Appli
25	119	4.6	1765	15	US-10-037-182-8	Sequence 8, Appli
26	119	4.6	1786	9	US-09-873-676-113	Sequence 113, Appli
27	119	4.6	1786	10	US-09-938-275-6	Sequence 6, Appli
28	119	4.6	1786	15	US-10-037-182-6	Sequence 23, Appli
29	118	4.5	928	12	US-10-080-608A-23	Sequence 23, Appli
30	118	4.5	928	12	US-10-370-685-112	Sequence 112, App
31	118	4.5	1289	10	US-09-712-363-259	Sequence 259, App
32	118	4.5	2285	10	US-09-932-183A-2	Sequence 2, Appli
33	117.5	4.5	931	12	US-10-032-585-7600	Sequence 7600, Ap
34	117.5	4.5	1192	12	US-10-053-662A-32	Sequence 32, Appli
35	117.5	4.5	1441	15	US-10-128-714-8123	Sequence 8123, Ap
36	117.5	4.5	4999	10	US-09-976-059-14	Sequence 14, Appli
37	117	4.5	969	15	US-10-078-531-8	Sequence 8, Appli
38	117	4.5	971	15	US-10-078-531-7	Sequence 7, Appli
39	117	4.5	1008	15	US-10-078-531-2	Sequence 2, Appli
40	116.5	4.5	1388	15	US-10-146-473-82	Sequence 82, Appli
41	116	4.5	971	15	US-10-078-531-6	Sequence 6, Appli
42	116	4.5	974	15	US-10-156-761-10499	Sequence 10499, A
43	116	4.5	1039	15	US-10-128-714-3123	Sequence 3123, Ap
44	116	4.5	1441	12	US-10-412-897-3	Sequence 3, Appli
45	116	4.5	1641	14	US-10-017-216-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-09-738-626-3970  
; Sequence 3970, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIALI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patentin ver. 3.0  
; SEQ ID NO 3970  
; LENGTH: 463  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3970

Query Match 18.6%; Score 483.5; DB 10; Length 463;  
Best Local Similarity 27.7%; Pred. No. 1.1e-31;  
Matches 124; Conservative 102; Mismatches 186; Indels 35; Gaps 9;  
QY 81 TIAVIGLSVHTAPVEMREKLAVALPRAIQELTSLNHIEEAVALSTCNRMIEIYVVALS 140

Db 2 SVLIWGMHRSAPVSELLERLSMDDSVRGTTQALLGRASLSALIVSTCNRELEVYVTSS 61  
Qy 141 WNRGIREVYVDMWMSKSGIPASERLREHLFILRSSDATRHLFEVSAGLDSVLGEGQILAQV 200  
Db 62 FHTGVNDVVEVLHEASGVVDIETLRCGLVYRYADAAAEHMLVVTSGLDNVLGEGQILQV 121  
Qy 201 KOVVRSGQNSGGLGNIDRMFKDAITAGKRVRSSTNIGSAGVSVSSAAVELALMKL---P 257  
Db 122 RTAYQAAEYSGVGPALHSLTQTALHTGKRVHSETAIDDAGASMYSFADRALVQMGDLS 181  
Qy 258 KSEA--LSAR--MLLIGAGKMGKLVIKHLVAKGCKVWVNRVSVERVDAIREEMKDIEIVVR 315  
Db 182 EAEAPLSKTAIVLACAGAMSSLAATHLGRAGISNLIIMANRTLERAEHLAEHSLGAVPAE 241  
Qy 316 PLSDMYQAAA--EADVTTSTASSETSLFAKEHAELPPVSDTMGGVRLFVDISVPRNVSA 373  
Db 242 VVEYDQASAVNRVDLVVSATGDDF-----TVKPEDIPEGASLMVLVDSMPERDID 293  
Qy 374 CVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETPTTIKK 433  
Db 294 ACADLPVGVDLVNIERLHKASREGSGMAPSEEEALAVREELDSFTSQRINDIVPAVSA 353  
Qy 434 LRSYADRIASELEKCLQKV-----GEDALTKMRRRAIEELSTGIWNKLLHGLQL 485  
Db 354 LRRQAASVGSDELRLRQAPGISEVWGE--VEKTVR-----VVDKLLHEPTVVR 403  
Qy 486 R--CDGSDSRTLDETLNHNHMRMFS 510  
Db 404 KELAARSGSISYDSALQELFGLLESAS 430

## RESULT 2

US-09-882-227-508  
; Sequence 508, Application US/0988227  
; Publication No. US20030158396A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleanthous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Ooomen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the  
; TITLE OF INVENTION: Genome  
; FILE REFERENCE: 06132/047002  
; CURRENT FILING DATE: US/09/882,227  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 08/902,615  
; PRIOR FILING DATE: 1997-07-29  
; NUMBER OF SEQ ID NOS: 638  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 508  
; LENGTH: 449  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-882-227-508

Query Match 16.4%; Score 428.5; DB 12; Length 449;  
Best Local Similarity 28.8%; Pred. No. 3.7e-27;  
Matches 127; Conservative 81; Mismatches 186; Indels 47; Gaps 9;  
Qy 94 VEMREKLAVABELPRAIQEL--TSLNH--IEEAAVLSTCNRMIEYVW--ALSWNRGI 145  
Db 22 LEMREKLAIN--ATLKEFLQTIKNCPNKECWLSTCNREIYVSLKHGANTNEQ 78  
Qy 146 REVVDWMSKSGIPASERLREHLFILRSSDATRHLFEVSAGLDSVLGEGQILAQVQVVR 205  
Db 79 NALLKILAQNKMVSVDLEKCVLMNTDSEAVHHVFSVCSSLDLSLVGQTITGQMKAYK 138  
Qy 206 SQNSGGLGNIDRMFKDAITAGKRVRSSTNIGSAGVSVSSAAVELALMKLPKSEALSAR 265  
Db 139 FAFEEKCSKDLTLLHFAKCAAKVRNLTGISQGVSSISSVAKEALNIFEKERIKDK 198

Qy 266 MLLIGAGKMGKLVIKHLVAKGCKVWVNRVSVERVDAIREEMKDIEIVRPLSDMYQAAA 325  
Db 199 ALVIGLGEAQLVIGLHLLNKQPEALILGRNAKPEDFIKELEPKKVSFQNIENLWAYIN 258  
Qy 326 EADVVTSTAS-----ETSLFAKEHAELPPVSDTMGGVRLFVDISVPRNVSAV 375  
Db 259 EYELFCATSPHPFIVQNRMLKETIFR-----RPFEDLAVPRNIEKPV 301  
Qy 376 SEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETPTTIKK 435  
Db 302 --LONIPLYSVDDLEPMVRENVENRQSRMRAYEIVGLATMEFYQWIOSLEVEPIVKDL 359  
Qy 436 SYADRIASELEKCLQKVGEDALTKMRRRAIEELSTGIWNKLLHGLHPLQHLRCDGSDSRTL 495  
Db 360 ELARISAKELQKALK--RYVPKEVENNTEKILHNAFNTFLHNPTIALKKNQAQK---- 412  
Qy 496 DETLENMHALNRMFSLDMEKA 516  
Db 413 BESDVLVGAIKNLFNLDKNSA 433

## RESULT 3

US-10-156-761-12273  
; Sequence 12273, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 12273  
; LENGTH: 569  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-12273

Query Match 16.4%; Score 427; DB 15; Length 569;  
Best Local Similarity 24.4%; Pred. No. 7.1e-27;  
Matches 137; Conservative 94; Mismatches 184; Indels 146; Gaps 14;  
Qy 81 TIAVIGLSVHTAPVEMREKLAVABELPRAIQELTSLNHIEEAAVLSTCNRMIEYVVALS 140  
Db 2 SLLVVLGSHRSAPVSVLERATLTADAQVKKLQOTVAABEPATEAAVATCNRIELYADV 61  
Qy 141 WNRGIREVYVDMWMSKSGIPASERLREHLFILRSSDATRHLFEVSAGLDSVLGEGQILAQV 200  
Db 62 FHAGVAELSTLLAQHSGVGLLEELTPLYVHVYEDRAVHHLFSVACGLDSMVVGEQILQOI 121  
Qy 201 KOVVRSGQNSGGLGNIDRMFKDAITAGKRVRSSTNIGSAGVSVSSAAVELALMKLPKSE 260  
Db 122 KDALATAQELHTAGRLNLDLFQOALRTGKRAHSETGIDRAQSL-----VTFGLEQLSAGT 177  
Qy 261 ALSA-----RMLLIGAGKMGKLVIKHLVAKGCKVWVNRVS--VERVDAIREEMKDIEI 312  
Db 178 AVEAWKGRKALVIGGMSLSLAATLARAGVAEVIANKRTPDRAERLAQILTEGDDTDV 237  
Qy 313 VYR--PLSDMYQAAAEEADVVTSTAS-----ETSLEA-- 342  
Db 238 LARAVPMDAVPAELTRADVAVSCTGATGLVLTAEAAVAEEGRTGTPVAVREETPASAG 297

```

QY 343 -----KEH----- 345
Db 298 GLAPAGTDEGCPLDLSAVOGATGFSVMGEAAVAGMDAATLEQAAAWDRGTVDRRDSRT 357
QY 346 -----AEALPPVSDTM-GVRL-----FVDSVPRNVSACVSEYGA 380
Db 358 PEVEAELITALAATAVAGRLPERRPBPVVEAPRAPALALLDLAMPDRIDAAVHRLG 417
QY 381 ARVNVDDLKEVVE-----ANKEDRLRKAMEAQTIIITEELRFEAWRDSLETPVTKLRS 436
Db 418 VRLVDIESLAASADAPWAADVLR-----IVADEVAAFGAQRAAHITPTVVALRT 471
QY 437 YADRIRASELCKLOKQGEDALTCKMRRRAIBELSTGIVNKLHGP-----LOHLRCDGSDSR 493
Db 472 MAADVANEIARLDGRL-----PGLDEKQGEITQTVRVVDKLLHAPTAVRVKQLAAEPGGAG 529
QY 494 TLDETLENNHMRMFSLDME 514
Db 530 YAD-----ALRTLFDLDPE 543

```

## RESULT 4

US-10-156-761-10395

; Sequence 10395, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156.761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 45109

; SEQ ID NO 10395

; LENGTH: 1258

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-10395

## Query Match

Best Local Similarity 5.5%; Score 143; DB 15; Length 1258;

Matches 124; Conservative 76; Mismatches 214; Indels 172; Gaps 24;

```

QY 45 DAAGVEAQAQAVAKAASVAA-----LEQPKISADRYMKERSTIAV-IGLSVHTAP 93
Db 43 ESADLGYAEOLENNAGIQADQLRADAEELSQAQOTRIQEHAEQAARLQAEHTEA 102
QY 94 VENEKRL-----AVABELWPAIQELTSLNHIEE-----AAV 125
Db 103 VTRRQQLDQLAERBQTVESHVNVENVAQLRASEQQARRL--VEESRAEABQALAA 160
QY 126 LSTCNRM-----EIVVVALSWNRGIREVVDMWSKSGIPASELRHFLTIRS 172
Db 161 RABAEERIAABARQRLNTDAESARAEAEALLRRARTDAERLLNASTQAQEAHTAEQLRS 220
QY 173 SDATRHLPFVSAGLDSLVIGEGQILAQVKQVRSQNSGGLGKNDIMFKDAITAGKVR 232
Db 221 STAT-----ESDA-----ARQASELSRTAEQRMTEADAALREAR 255
QY 233 SETNISSGAVSVSSAAVELMLPKSEALSARMLLIGAGKMGKLVTKHIVAKGCKVVV 292
Db 256 AE-----ADKVLTEAKEAAKALASAESANEQRT-----RTAKEQVARLVSEATKEAT 304
QY 293 VNRSEVRVDAIREMKDIEIVYRPLSDMYQAAAEADVFTSTASET-SLFAKEHAELPP 351

```

```

Db 305 TKAAEEVVA-----DAKAAEKILTAAADKARSLTAETASQLTK 345
QY 352 VSDTMGGVRLFVDISVPRNVSACVSEYGAARVYVDDLKVEVYANKK-DRLR----- 402
Db 346 AARTAEADVNLKASBEAKSTTKAAATEE--ABRIRR-----EAETRADRLRAEAHDIAE 395
QY 403 -----KAMAAQTI-ITEELRRFEAWRDSL--ETVPTIKKLRSADRIRASELEKC 449
Db 396 QLKGTAKDDTKYRAKTVELQSEARRLRGDAEQLRADAAVEGERIRSEARREAVQOIIEBA 455
QY 450 LQKVGEHDALTK-----KMR-RAIEELSTGIVNKLHGPQLHRLCDGSD 491
Db 456 -AKTAEELLAKAKADADELRTAATTDSEKVTETAIERATT--LRQABETLRTNEAR 512
QY 492 SRTLDETLENNHMLN-----RMFSLDMEKAIIEQKIKAKVEKTQ 530
Db 513 HRA--EAVEQSEAIKEDAERAARELRDDSERAIERARAEANDELTR 556

```

## RESULT 5

US-10-080-608A-19

; Sequence 19, Application US/10080608A

; Publication No. US20030198956A1

; GENERAL INFORMATION:

; APPLICANT: Makowski, Lee

; APPLICANT: Hyman, Paul

; APPLICANT: Williams, Mark

; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES

; FILE REFERENCE: 8471-010-999

; CURRENT APPLICATION NUMBER: US/10/080.608A

; CURRENT FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 180

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19

; LENGTH: 975

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-10-080-608A-19

## Query Match

Best Local Similarity 5.4%; Score 139.5; DB 12; Length 975;

Matches 117; Conservative 104; Mismatches 230; Indels 129; Gaps 22;

```

QY 2 ATTTSATTAATAAAATTAKPRGSSAL-----CORVAGGRRRSVVVRCDAAGVEA 51
Db 405 AACTATAAAALAAQRTALANNSASVAVNEQARLATECERL-----YQQLDDKDEE 455
QY 52 QAQA-VAKAASVAALAEQFKISAD---RYMKERSTIAVIGLSVHTAPVEMREKLAVAEEL- 106
Db 456 NQDSQYAEQLKEQVMEQEELIANARREYETLQSEMARIQENESAKSEVKEVLQALELA 515
QY 107 --WPRATQELTSLNHIEEAAVLSTCNRMETVVALSWNRGIREVVDMWSKSGIPASELR 164
Db 516 VNYDQKSOEIDNKKNDIDALNEBLQKQSVFNAASTELQKDKMSSHQKRR-----ITEML 571
QY 165 EHLFILRSSDATHLPFVSAGLDSLVIGEGQILAQVKQVRSQNSGGLGKRN-----I 217
Db 572 TNL--RDLEGEV--QAIAPGESSI--DLQMSALAGTDAKVEDFTMARLFI 618
QY 218 DRMFKDAITAGKVRVS-ETNISSGAVSVSSAAVELALMLPKSEALSARMLLIGAGKMGK 276
Db 619 SKMTEAKNTAQCSNMETQQADSNKKISYEKDLGEYRL----- 658
QY 277 LVIKHLVAKGCKKVVVNRSEVRVDAIREEMKDIEIVYRPLSDMYQAAAE--ADVFTST 334
Db 659 -----LISQHEARKMLQBSMRBAENKKTLEBQIDSLRECEAKLKAHEH 703
QY 335 ASSTSLFAKEHAELPPVSDT-MGVRLFVDISVPRNVSACVSEYGAARVYVDDLKVE- 392
Db 704 VSAVNAEERQABEELSMFDSQMDLR-----EAHTQVSELRDEI-AAKHQHEMDKMDVH 758
QY 393 -----VEANKEDRLRKAMEAQTIIITEELRRFEAWRD-----SLETVP 429

```

```
Db 759 QKLLAHQOMTADYKQSDAEKSELQNIILTNERREQARKOLKGLDVTAKELQTLH 818
Qy 430 TIKKL--RSYADRIASELCKLQKVGEDALTQKMRRAIEELSTGI-----VNKLLHGPL 482
Db 819 NLRKLFVQDLQQRIRKNVNV--BSEEDGGSLAQKQKISFLENNLDQLTKVHKQLVRDN 875
Qy 483 QHLRCD-GSDSRTLDETLENHMLNRMPSLDMKAEKAIIEQK 521
Db 876 ADLRCELPLKRLRCTMRVKALETALKEAKGAMRDRK 915

RESULT 6
US-10-370-685-108
; Sequence 108, Application US/10370685
; Publication No. US20030215903A1
; GENERAL INFORMATION:
; APPLICANT: Hyman, Paul
; APPLICANT: Goldberg, Edward
; TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional Elements
; FILE REFERENCE: NANF-P-004
; CURRENT APPLICATION NUMBER: US/10/370,685
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 10/080,608
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 108
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-370-685-108

Query Match 5.4%; Score 139.5; DB 12; Length 975;
Best Local Similarity 20.2%; Pred. No. 0.012;
Matches 117; Conservative 104; Mismatches 230; Indels 129; Gaps 22;

Qy 2 ATTTATTAAAAAATKPRGSSAL-----CORVAGGRRRRSGVVRCDAAAGVEA 51
Db 405 AAQTAAAEALAAQRTALANWSASVAVNEQARLATECERL-----YQOLDKDBEI 455
Qy 52 QAQA-VAKAASVAALAEQPKISAD---RYMKERSTIAVIGLSVHTAPVEMREKLAVAEEL- 106
Db 456 NQOQOYAEQLKEQVMEQBELIANARREYETLQSEMARIQENESAKEVEVLQALEELA 515
Qy 107 --WPAIAGELTSLNHEAAVLSTCNRMIEYVVALSWNRGIREVVDMWKKSGIPASBLR 164
Db 516 VNYDQKSOEIDNKNKDIDALNEELQOKQSVFNAASTELQQLKDMSSHOKKR-----ITEML 571
Qy 165 EHLFILRSSDATRHLFEYSAGLDSLVLGEGQILAQVKQVRSGONGSLGKN-----I 217
Db 572 TNLL-----ROLGEVG--QAATAPGESSI--DLKMSALAGTDAKVEEDFTWARLFI 618
Qy 218 DRMFKDAITAGKVRVS-ETNISSGAVSVSSAAVELALMKLPKSEALSARMLLIGAGKMGK 276
Db 619 SKMKTEAKNIAQRCSNMETQOADSNNKISEYKDLGEVRL----- 658
Qy 277 LVIRHLVAKGCKVNVNRSVERVDAIREEMKDIEIVRPLSDMYQAAAE--ADVFTST 334
Db 659 -----LISQHEARMKSLQESMREANKRTLEEQIDSLREECALKAHEH 703
Qy 335 ASETSLFAKEHAELPPVSDT-MGGVRLFVDISVPRNVSVACVSEVGAARVNVDDLKEV- 392
Db 704 VSAVNAEQRABEELRSNMFDSQMDLRL---EAHTRQVSELDRDI-AAKHHEMDEMVDVH 758
Qy 393 -----VEANKEDRLRKAMEAQTIITEELRRFEAWRD-----SLETVP 429
Db 759 QKLLAHQOMTADYKVRQEDAEKSELQNIILTNERREQARKOLKGLDVTAKELQTLH 818
Qy 430 TIKKL--RSYADRIASELCKLQKVGEDALTQKMRRAIEELSTGI-----VNKLLHGPL 482
Db 819 NLRKLFVQDLQQRIRKNVNV--BSEEDGGSLAQKQKISFLENNLDQLTKVHKQLVRDN 875
```

```
Qy 483 QHLRCD-GSDSRTLDETLENHMLNRMPSLDMKAEKAIIEQK 521
Db 876 ADLRCELPLKRLRCTMRVKALETALKEAKGAMRDRK 915

RESULT 7
US-10-037-182-12
; Sequence 12, Application US/10037182
; Publication No. US20030044899A1
; GENERAL INFORMATION:
; APPLICANT: Trygsvaen, Karl
; APPLICANT: Doi, Masayuki
; APPLICANT: Thyboll, Jill
; TITLE OF INVENTION: Recombinant Laminin 10
; FILE REFERENCE: 99-274-F
; CURRENT APPLICATION NUMBER: US/10/037,182
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,449
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/279,282
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1725
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-037-182-12

Query Match 5.3%; Score 138; DB 15; Length 1725;
Best Local Similarity 19.0%; Pred. No. 0.038;
Matches 125; Conservative 101; Mismatches 243; Indels 188; Gaps 27;

Qy 28 QRVAG-GGRRS-----GVRCDAAAGVEAQA--QAVAKAASVAALAEQPKI 70
Db 1042 CQCPGFGGRTSCQELFWGDPDVECRACDCDPRGISTPQCDSTGQCVCVEGPRC 1101
Qy 71 SADRYMKERSTIAVIGLSVHTAPVEMREKLAVAELMPRAIQELTSLNH--IIEAAVJST 128
Db 1102 -----DKCTRGYGVFPDCTPCHQ-----CFALMDAIIIGELTNRTHKFEKAKALKI 1148
Qy 129 CNRMIEYVVALSWNRGIREVVDMWKK-----SGIPASELREHLFIL-----RSSD 174
Db 1149 SGVIGPY-----RETVDSVEKKVNEIKDIAQSPAAEPKUNYIGILFEAEKLTND 1198
Qy 175 ATRHLFEVSAGLDSL-----LGEGQI-----LAQVKQVRSGONGSLG 214
Db 1199 VTEQMAQVEVKLTDTASQSNSTAGELGALQAAEASLDTKVKELAEQLEFIKNSDIQGAL- 1257
Qy 215 KNIDRMFKDAITAGKVRVSETNIISSGAVSVSSAA--VELALMK-----LPKSEALSARM 267
Db 1258 DSITKYPQMSLEAEKRVNASTTDPNSTVQOSALTDRVEDLMLERESPFFKEQBEQARLL 1317
Qy 268 LIGAGKMGKLV-----KHLVAKGCKVNVN 295
Db 1318 DELAGKIQSLDLSAAQMTGTPPGACSESECGPNCRTDEGEKKCGGCGGLVTVAH 1377
Qy 296 SV--ERVDAIRE--EMKDIEIVRPLSDMYQAAAE-----DYVFTSTASSETSL----- 340
Db 1378 SAMQKAMDPRDVLASALAEVQLSKMYSEAKVRADEAKQNAQDVLTKTNATKEKVDKSNE 1437
Qy 341 -----FAKEHAELPPVSDTMGGVRLFVDISVPRNVSVACVSEVGAARVNVDDL 389
Db 1438 DLRLIKQIRNFLTDSADLSIEAVANVFLKSGNASTPQOLQNLTDI--RERVETLSQV 1496
Qy 390 KEVVEANKEDRLRKAMEAQTIITEELRRFEAWRD-----SLETVP 429
Db 1497 EVILQQAADIAR-----AELLLEAKKASKATDKVKTADMKVKEALEEAKQAQVA-AEKA 1551
Qy 450 LQKVGED-----ALT-----KMRRAIEEL-----STGIV 474
Db 1552 IKQADEDIQGTQNLTSIETAASEBTLTNASQRIKLERNVLEELKRAAQNSEAEYI 1611
```



QY 475 NKLHGPLQHLRCDGSDRTLDTELNNHMLNRMFSLDMEKAIIEQKIKAKVEKTON 531  
Db 1612 EKVVSVKQNA---DDVKLTLDGELDEKY--KKVESLIAQKTEESADARRKAEILLQ 1663

RESULT 8  
US-09-938-275-7  
; Sequence 7, Application US/09938275  
; Patent No. US20020111309A1  
; GENERAL INFORMATION:  
; APPLICANT: Gerardo Castillo  
; APPLICANT: Alan Snow  
; TITLE OF INVENTION: Therapeutic and Diagnostic Applications  
; FILE REFERENCE: PROTEO.P03  
; CURRENT APPLICATION NUMBER: US/09/938,275  
; CURRENT FILING DATE: 2001-08-16  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 1786  
; TYPE: PRT  
; ORGANISM: Mus Musculus  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: Swissprot P02469  
; DATABASE ENTRY DATE: 1989-07-01  
US-09-938-275-7

Query Match 5.3%; Score 138; DB 10; Length 1786;  
Best Local Similarity 19.0%; Pred. No. 0.04;  
Matches 125; Conservative 101; Mismatches 243; Indels 188; Gaps 27;

QY 28 CORVAG-GRRRS-----GVRCDAGVEAQA--QAVAKAASVAALAEQPKI 70  
Db 1103 CQCMFGGRTCECOELFWGDPDVECRACDPRGIETPCDQSTGQCVCVGVGPRC 1162

QY 71 SADRYMKERSTIAVIGLSVHTAPVEMREKLAVAEELWPRAIQELTSLNH--IEEAALVST 128  
Db 1163 -----DKCTRGYSVFPDCTPCHQ-----CFALMDAIIIGELNTRTHKFLKAKAKI 1209

QY 129 CNRMEIYVVALSWNRGIREVVDMMSKK-----SGIPASELRHLPIL-----RSSD 174  
Db 1210 SGVIGPY-----RETVDSEKKNWEIKDILAQSPAAEPLKNIGILFEEAEKTKD 1259

QY 175 ATRHLFEVSAGLDSL-----LGEGQI-----LAQVKQVRSQNSGGIG 214  
Db 1260 VTERKMAQVEVKLTDTASQSNSTAGELGALQAEASLDKTVKELAEQLFKNSDIQGAL- 1318

QY 215 KNIDRMFKDAITAGKRVRSSETNISGAVSVSSAA---VELALMK---LPKSEALSARM 267  
Db 1319 DSITKYFQMSLEAEKRVNASITDPNSTVEQSALTDRVEDLMLERESPFKEQOEQARLL 1378

QY 268 LIGAGKMGKLV-----KHLVAKGCKKVVVNR 295  
Db 1379 DELAGKQLSLDLSAAQMTCTPGADCSSECCGPNCRBTDEGEKKGCGPGCGGLVTVAH 1438

QY 296 SV--ERVDATRE---EMKDIIVYRPLSDMYQAAAEA-----DVVFTTASSETS- 340  
Db 1439 SAWQKAMDFDRDVLALAEVQLSKWVSEAKVRADKQNAQDVLTKTNATKEKVDKXNE 1498

QY 341 -----FAKEHAELPPVSDTMGGVRLFVDISVPRNVSVACVSGVGAARVNVVDL 389  
Db 1499 DLRLNIKQIRNFLTSDSADLSIEAVANEVLKSGNASTPQQLNLTEDI--RERVETLSQV 1557

QY 390 KEVVEANKEDRLRKAMEAQTIITEELRPFPAWRDSLETPTIKKLSYADRIRASELEKC 449  
Db 1558 EVILQQAADIAR-----AELLLEAKRASKSATDKVTADVMVKEALEBAEKAQVA-AEKA 1612

QY 450 LQKVGED-----ALT-----KQMRRAIEEL-----STGIV 474  
Db 1613 IKQADEIQGTQNLTLTSEETAASEETLTNASQRIKSLERNVEELKRAQNSGEAEYI 1672

QY 475 NKLHGPLQHLRCDGSDRTLDTELNNHMLNRMFSLDMEKAIIEQKIKAKVEKTON 531

Db 1673 EKVVSVKQNA---DDVKLTLDGELDEKY--KKVESLIAQKTEESADARRKAEILLQ 1724

RESULT 9  
US-10-037-182-10  
; Sequence 10, Application US/10037182  
; Publication No. US20030044899A1  
; GENERAL INFORMATION:  
; APPLICANT: Trygvsason, Karl  
; APPLICANT: Doi, Masayuki  
; APPLICANT: Thyboll, Jill  
; TITLE OF INVENTION: Recombinant Laminin 10  
; FILE REFERENCE: 99-274-F  
; CURRENT APPLICATION NUMBER: US/10/037,182  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,449  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/279,282  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 1786  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-037-182-10

Query Match 5.3%; Score 138; DB 15; Length 1786;  
Best Local Similarity 19.0%; Pred. No. 0.04;  
Matches 125; Conservative 101; Mismatches 243; Indels 188; Gaps 27;

QY 28 CORVAG-GRRRS-----GVRCDAGVEAQA--QAVAKAASVAALAEQPKI 70  
Db 1103 CQCMFGGRTCECOELFWGDPDVECRACDPRGIETPCDQSTGQCVCVGVGPRC 1162

QY 71 SADRYMKERSTIAVIGLSVHTAPVEMREKLAVAEELWPRAIQELTSLNH--IEEAALVST 128  
Db 1163 -----DKCTRGYSVFPDCTPCHQ-----CFALMDAIIIGELNTRTHKFLKAKAKI 1209

QY 129 CNRMEIYVVALSWNRGIREVVDMMSKK-----SGIPASELRHLPIL-----RSSD 174  
Db 1210 SGVIGPY-----RETVDSEKKNWEIKDILAQSPAAEPLKNIGILFEEAEKTKD 1259

QY 175 ATRHLFEVSAGLDSL-----LGEGQI-----LAQVKQVRSQNSGGIG 214  
Db 1260 VTERKMAQVEVKLTDTASQSNSTAGELGALQAEASLDKTVKELAEQLFKNSDIQGAL- 1318

QY 215 KNIDRMFKDAITAGKRVRSSETNISGAVSVSSAA---VELALMK---LPKSEALSARM 267  
Db 1319 DSITKYFQMSLEAEKRVNASITDPNSTVEQSALTDRVEDLMLERESPFKEQOEQARLL 1378

QY 268 LIGAGKMGKLV-----KHLVAKGCKKVVVNR 295  
Db 1379 DELAGKQLSLDLSAAQMTCTPGADCSSECCGPNCRBTDEGEKKGCGPGCGGLVTVAH 1438

QY 296 SV--ERVDATRE---EMKDIIVYRPLSDMYQAAAEA-----DVVFTTASSETS- 340  
Db 1439 SAWQKAMDFDRDVLALAEVQLSKWVSEAKVRADKQNAQDVLTKTNATKEKVDKXNE 1498

QY 341 -----FAKEHAELPPVSDTMGGVRLFVDISVPRNVSVACVSGVGAARVNVVDL 389  
Db 1499 DLRLNIKQIRNFLTSDSADLSIEAVANEVLKSGNASTPQQLNLTEDI--RERVETLSQV 1557

QY 390 KEVVEANKEDRLRKAMEAQTIITEELRPFPAWRDSLETPTIKKLSYADRIRASELEKC 449  
Db 1558 EVILQQAADIAR-----AELLLEAKRASKSATDKVTADVMVKEALEBAEKAQVA-AEKA 1612

QY 450 LQKVGED-----ALT-----KQMRRAIEEL-----STGIV 474  
Db 1613 IKQADEIQGTQNLTLTSEETAASEETLTNASQRIKSLERNVEELKRAQNSGEAEYI 1672

QY 475 NKLHGPLQHLRCDGSDRTLDTELNNHMLNRMFSLDMEKAIIEQKIKAKVEKTON 531





```
RESULT 14
US-09-815-242-5230
; Sequence 5230, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5230
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5230

Query Match          4.9%; Score 128; DB 9; Length 869;
Best Local Similarity 21.6%; Pred. No. 0.094;
Matches 119; Conservative 85; Mismatches 197; Indels 150; Gaps 28;

Qy 47 AGVEAQ-----AQAVAKASVAALAEQFKISADRYNKE-----RSTIAVIGLSVH 90
Db 72 ASVEGDNIQYQYISQAN-----QLITKAESYKMEYDEYISMEHILRSAMDIDQTTKH 126
Qy 91 TAPVEMREKLAVAEELWPRAIQELTSLNHIEEAVALSTCNRMETVYVVALSWNRGIREVVD 150
Db 127 -----YINKKVEIKDI-----IKKVRGGNHV-----TSQNPEVNYEALA--KYGRDLVE 169
Qy 151 WMSKKSQIPA-----SELREHLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQVKQVVR 206
Db 170 EVRQGMDFVIGRDEEIRNTIRILSRKTKNPNVLIGEPGCKTAIVEG--LAQ--RIVKK 225
Qy 207 GQSGGLKNDIRMFKDAITAGKVRSETNISSGAVSVSSAAVELALMKLPKSEA-----261
Db 226 DVPESLDKTVFELDLSALVAGAKYRGEFERLKAV-----LKEVKESDGRILF 275
Qy 262 LSARMLLIGAGKM--GKL-----VIKHLVAKG-----CKKVVVNRVERV 300
Db 276 IDEIHLVAGKTDGADGNWMLKPLMARGELHCIGATTLINEYREYIEKOSALERRFOKV 335
Qy 301 DAIREMKDIEIVRPISDMYQ-----AAAAADVFTSTASSETSLFAKEHA--346
Db 336 AVSEPDVEDTISILRGLKERYEVHGVRIQDRALVAAAE-----LSDRYITDRLPKAI 390
Qy 347 -----EALPPVSDTMGGVRLFDVDSVPRNV-----SACVSEVCAARVYNVDLKEVVEA 395
Db 391 DLVDQACATRTENGSGNPTELD--QVNRVRVMOLEIESALKNESDNASKORLQELQEEAN 449
Qy 396 NKE-----DLRLKAMEAQTIITEELRRFEAWRDSLETPTIKLRSVADRIRAS--ELE 447

Db 450 EKEKQAAALQSRVSESEKIANLQEKRAQLDESROALBDAQTNNNLEKAAELQYGTIPQLE 509
Qy 448 KCLQKV-----GEDALTCKMRAI---EELS-----TGIV-----NKLLH-G 480
Db 510 KELRELEDNPFQDEQED--TDRMIREVVUTDEIGDIVSQWTGIPVSKLVETEREKLLHLS 567
Qy 481 PLOHLRCDGSD 491
Db 568 DILHKRVVQGD 578

RESULT 15
US-09-815-242-12526
; Sequence 12526, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12526
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12526

Query Match          4.9%; Score 128; DB 9; Length 882;
Best Local Similarity 21.6%; Pred. No. 0.096;
Matches 119; Conservative 85; Mismatches 197; Indels 150; Gaps 28;

Qy 47 AGVEAQ-----AQAVAKASVAALAEQFKISADRYNKE-----RSTIAVIGLSVH 90
Db 85 ASVEGDNIQYQYISQAN-----QLITKAESYKMEYDEYISMEHILRSAMDIDQTTKH 139
Qy 91 TAPVEMREKLAVAEELWPRAIQELTSLNHIEEAVALSTCNRMETVYVVALSWNRGIREVVD 150
Db 140 -----YINKKVEIKDI-----IKKVRGGNHV-----TSQNPEVNYEALA--KYGRDLVE 182
Qy 151 WMSKKSQIPA-----SELREHLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQVKQVVR 206
Db 183 EVRQGMDFVIGRDEEIRNTIRILSRKTKNPNVLIGEPGCKTAIVEG--LAQ--RIVKK 238
Qy 207 GQSGGLKNDIRMFKDAITAGKVRSETNISSGAVSVSSAAVELALMKLPKSEA-----261
Db 239 DVPESLDKTVFELDLSALVAGAKYRGEFERLKAV-----LKEVKESDGRILF 288
Qy 262 LSARMLLIGAGKM--GKL-----VIKHLVAKG-----CKKVVVNRVERV 300
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 12:53:43 ; Search time 21 Seconds

(without alignments)  
1669.860 Million cell updates/sec

Title: US-10-018-902-4

Perfect score: 2606

Sequence: 1 MATTTATTAAAAAATAKPP.....DMEKALIEQIKAKVETQN 531

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	557.5	21.4	435	4	US-09-252-991A-30562
2	543.5	20.9	448	4	US-09-134-001C-3055
3	498	19.1	429	4	US-09-328-352-4643
4	206.5	7.9	341	4	US-09-198-452A-759
5	139.5	5.4	975	4	US-09-914-259-19
6	132.5	5.1	871	4	US-09-134-001C-3979
7	121	4.6	1939	4	US-09-310-187A-1
8	119.5	4.6	1886	4	US-08-938-105-3
9	119	4.6	1196	1	US-08-144-121-4
10	119	4.6	1196	2	US-08-735-893-4
11	118	4.5	928	4	US-09-914-259-23
12	118	4.5	2285	4	US-09-308-375-2
13	117	4.5	573	4	US-09-252-991A-18744
14	116.5	4.5	1388	4	US-09-572-191-2
15	116.5	4.5	1388	4	US-09-723-262-2
16	116.5	4.5	1388	4	US-09-723-219-2
17	115.5	4.4	3248	1	US-08-353-700-1
18	115.5	4.4	3248	5	PCT-US95-16216-1
19	115	4.4	2101	1	US-08-466-390-4
20	115	4.4	2101	1	US-08-470-950-4
21	115	4.4	2101	1	US-08-467-781-4
22	115	4.4	2101	1	US-08-195-487-4
23	115	4.4	2101	2	US-08-483-924-4
24	115	4.4	2101	3	US-09-452-294-1
25	115	4.4	2101	5	PCT-US93-06160-4
26	114.5	4.4	439	4	US-09-252-991A-31097
27	114.5	4.4	941	4	US-09-336-447A-9

28	113.5	4.4	2482	1	US-08-328-254-6	Sequence 6, Appli
29	113	4.3	756	3	US-09-085-199B-9	Sequence 9, Appli
30	112.5	4.3	614	4	US-09-252-991A-31412	Sequence 31412, A
31	112	4.3	748	4	US-09-252-991A-31491	Sequence 31491, A
32	112	4.3	1713	3	US-08-600-982-24	Sequence 24, Appl
33	112	4.3	1713	5	PCT-US94-10261A-24	Sequence 24, Appl
34	111.5	4.3	592	2	US-08-736-770-6	Sequence 6, Appli
35	111.5	4.3	592	4	US-09-702-705-1809	Sequence 1809, Ap
36	111.5	4.3	592	4	US-09-736-457-1809	Sequence 1809, Ap
37	111.5	4.3	5215	3	US-09-105-537-2	Sequence 2, Appli
38	111	4.3	388	4	US-09-252-991A-21907	Sequence 21907, A
39	111	4.3	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
40	110.5	4.2	696	4	US-09-252-991A-16965	Sequence 16965, A
41	110	4.2	332	4	US-03-252-991A-29420	Sequence 29420, A
42	110	4.2	477	4	US-09-328-352-5428	Sequence 5428, Ap
43	110	4.2	2188	4	US-09-328-352-7763	Sequence 7763, Ap
44	109.5	4.2	467	4	US-09-252-991A-21258	Sequence 21258, A
45	109.5	4.2	1024	4	US-09-562-737-41	Sequence 41, Appl

## ALIGNMENTS

RESULT 1  
US-09-252-991A-30562  
; Sequence 30562, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30562  
; LENGTH: 435  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30562

Query Match		21.4%;	Score 557.5;	DB 4;	Length 435;	
Best Local Similarity		32.3%;	Pred. No. 2.3e-43;			
Matches		142;	Conservative	91;	Mismatches 187;	
			Indels	19;	Gaps 7;	
Qy	76	MKERSTTAVIGLSVH--TAPVEMREKLAVAEELWPRAIOELTSLNHIIEEAAVLSTCNRME	133			
Db	8	LAQRSCNFAIAGINHKHTASVAVRVAFTPEQMVEALQQLCRLTTSREAAILSTCNRS	67			
Qy	134	IYVVALLSWNRGIREVDWMSKSGIPASELREHLFILRSSDATRHLFEVSGAGLSVLGE	193			
Db	68	LYLEI--DHTADDVLAWLADYHRLTDELRACAVHQDEDAVHMMRVASGLDSMWLGE	125			
Qy	194	GQILAQVKQVVRSGQSGGLGKNDIRMFKDITAGKEVRSETNITSSGAVSVSSAAVELAL	253			
Db	126	PQLGQKMSAYAVAREAGTVGLLGRLFQATFTSTAKTVRTDTAIGENPVSVFAVSLA-	184			
Qy	254	MKLPKSEALSARMLLTGAGKMGKLVKHLVAKGCKVVRVNSVERVDATREEMKDIEIV	313			
Db	185	-KQIFSDLRHRSQALLGAGETITLVARHLPEQGVKRIVVANRTLERASLLAEQGAHAVL	243			
Qy	314	YRPLSDMYQAAARADVVTSTASETSLFAKEHAELAPPVSDTWGGVTLFVDISVPRNVA	373			
Db	244	---LSKEIPEELANSDIVISTASQLPILGKAVER--ALKQRKHKPMFMYDIAVPRDIEP	298			
Qy	374	CVSEVGAARYNVDDLKVEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETPTIKK	433			
Db	299	EVGELDDVLYSGVDDLHLHEVVAENLKSQCAQAAEELVGSVAFEFQRLRELAADVLR	358			



```

QY 128 TCNRMEIYVALSWNRGIREVVDWMSKSGIPASELREHL-----FILRSSDATRHLP 180
Db 52 TCHRAELY-----YYSEPIAQAAALLSELTSGGIRPYRHRGLSCFTHLP 96
QY 181 EYSAGLDSLVLGEGQILAQVKQVWRSGQNSGGLGKNIDRMFKDAITAGKRVSETHWISSG 240
Db 97 QVTSGLDSLIFGETEIQGVKRAYLKGSKERELPFDLHFLFOKALKEGKEYHSRIGFPDH 156
QY 241 AVSVSSAAVELALMKLPKSEALSARMLLIGAGMGKLVIKHLVAKGCKKVVVWNRVERV 300
Db 157 QVTIESVVQEI--LLSYDKS--IYTNFLFVGYSIDINRKAAYLYQHGHYRITFCSR----- 208
QY 301 DAIREEMKDIEIVYPLS-DMYQAAAEADVFTSTASETSLFAKEHAEALPVSDDTMGV 359
Db 209 -----QVVTAPYITLSRETLSPFQPDVIVFGSSESQSFDSLSCSLASIPK----- 256
QY 360 RLFDVDSIVPRNVSVACSEVGAARVY-NVDDLKEVVEANKEDRLRKAMEAQTIIT 412
Db 257 RIVDFDNVPR--TFLWKETPTGFVYLDIDFISECVQKRLOCTKEGVNKAALLT 308

RESULT 5
US-09-914-259-19
; Sequence 19, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-914-259-19

```

Query Match	5.4%;	Score 139.5;	DB 4;	Length 975;
Best Local Similarity	20.2%;	Pred. NO. 0.00062;		
Matches 117; Conservative 104; Mismatches 230; Indels 129; Gaps 22				
Qy	2	ATTTSATTAATAAATTAKPRGSSSAL-----CORVAGGGRRRSGVVRCDAAGVEA	51	
Dd	405	AACATAAAEAALAQAORTLANNASAVANNEQARLATTECERL-----YQQLDKDKDEI	455	
Qy	52	QAOA--VKAASAVALAEQPKISAD---RYMKERSTIAVLGLSVHTAPVMREKLAVAEEL-	106	
Dd	456	NQOQSOYAEOQLKEQVMEQBELIANARREYETLQSEMARIQENESAKEYEVQLQALEELA	515	
Qy	107	--WPAJQELTSLAHIEBAAVLSCTCNRMEEIVVALSWNRGIREVVUDWNMSKSGIPASELR	164	
Dd	516	VNYDQSQEIDNKVKNDIALNREELOQKOSVFNAASTELQQLKKDSSSHOKKR-----ITEML	571	
Qy	165	EHLFILRSSDATRHLEFVSAGLSLVLGEGQILAQVKOVVRSGNSGGIGKN-----I	217	
Dd	572	TNLL-----RDLEGVG---QAIPACESSI--DLKWSALAGTDASKVEEDFTMARLF	618	
Qy	218	DRMPKDAITAGKVRS-ETNWISSGAVSVSSAAVELALMKLKPSEALSARMILLIGAKMGK	276	
Dd	619	SKMKTAKNTAQCSNHMETQQADSNKKISEYEKGOLGEYRL-----	658	
Qy	277	LVIKHLVAKGCKKYVVNVNRSVERVDAREEMKDIEIVRPPLSDMYQAAA- -ADVVPFTST	334	
Dd	659	-----LISOHEARKMSIQESMRBANKKRTLEEQIDSILREECALKKAHH	703	
Qy	335	ASETSLPAKEHAERALPPVSDT-MCGVRLFDVIDSVPRNVSAVCSEVGAARVYNVDLKEV-	392	
Dd	704	VSVAHAEKQRABELSMFSOMDELN-----EAHTROVSELURDI-AAKHQEMDEMCDKH	758	

```

QY 393 -----VEANKEDRLRKAMEAQTIITEELRRFEAWRD-----SLETP 429
Db 759 OKLLLAHQMTADYKVRQEDA EKSSSELQNIILTNERREQAKDLKGLBETVAKELQTLH 818
QY 430 TIKL--RSYADRIASLEKCLKQVGEDALTKKMRRAIEELSTGI-----VNKLLHGPL 482
Db 819 NLRKLFVQDLOQRIRNVN---ESEEDEGGSLAQKQKISFLENNLDQLTKVHKQLVRDN 875
QY 483 QHLRCD-GSDSRTDLETLENNMHALNMFSLDMEKAIIEOK 521
Db 876 ADLRCELPKLEKLRCTMVERVKALETALKEAEGAWDRK 915

RESULT 6
US-09-134-001C-3979
; Sequence 3979, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3979
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3979

```

```

Query Match      5.1%; Score 132.5; DB 4; Length 871;
Best Local Similarity 21.5%; Pred. No. 0.0023;
Matches 123; Conservative 86; Mismatches 209; Indels 153; Gaps 26;

QY 52 QAQAVAKAAS-----VAALEQPKISADRYMKEST-----IAVIGLSVHTA---PVEM 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 32 EVEALLKAAMNENDSLFKSILERANIDVDQLIKAYDNQLSHYPTVQGDNVQYQOYISAKT 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 97 REKLAVABELWPRAIOELTSLNHTEEAVALS-----TCNMEIYVVALSWNRG----- 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 NELLDKAEKYMKSVEDEFISMEHLIRAIDTDTTQKWVGKNKVEIKEIITKVRGNNHVT 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 -----IREVDNMWKKSGIPA---SELREHLFILRSSDATRHLFEVSAGL 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 SQNEPVNVEALEKYGRLDVEVRQCKMDPVIGRDEEINRTIILSRKTKNNPVLIGEPGV 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 187 DSLVLGSGOILAQVQVRSGQSGGLKNDIRMPKDAITAGKRVSEITNISSGAVSSV 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 212 GKTAIVEG--LAQ--RIVRKDVPSLLDKTIFELDLSALVAGAKYGEFEERLKAV---- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 ANVELALMKLPKSEA-----LSARMLIGAGM--GKL-----VIKHLVAKG----- 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 -----LKEVKESSGRILFTIDEIHLVLGAGKTDCAMDAGNMLKPLMARGELHCIGATT 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 -----CKKVVVWVRSRVEDAIIEEMKDIEIVVRPLSDMYQ-----AAAEA 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 NEVREYIEKOSALERPFQKVGSPDVENTISILRGLKERYEVYHGVIQDRALVAAAE- 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 DVVFTSTASSETSLFAKEHA-----EALPPVSDTGGVRLFVDISVPRNV-----SACV 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 377 -----LSDRYITDRPLDPKALDLDVQACATIRTEGNSNPTELD-QVNRVMQLEIBESAK 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 376 SEVGAARVYNVDLKEVVEANKE-----DLRKAMEAQTIITBELRFEAWRDSLETVP 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 432 NESDNASKRLEBELQEELSNEKEQSSLKSRVEQEKEIAKVQEKRAELDSSRQALDAQ 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 430 TIKKLRSYADRIIRAS--ELBCKILO-----KVGEDALTCKOMRAI---EELS----- 470
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Db      492  TEGNLEKAAELOYGTIPQLEKELQEFEBAFQDETGED--SERMTREVVSDEIGDIVSQW 549
Qy      471  TGI-VNKLH-----GPLQHLURCDGSD 491
      ||| |||:
Db      550  TGIPSVKLAVETEBEKJLSLSDLHKRVVVGQD 580

RESULT 7
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; TITLE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; CURRENT FILING DATE: 1999-05-12
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match      4.6%; Score 121; DB 4; Length 1939;
Best Local Similarity 17.3%; Pred. No. 0.098;
Matches 106; Conservative 118; Mismatches 248; Indels 142; Gaps 22;

Qy      34  GRRRSGVVRCDAAGVEAQAQAAVAKAA--SVAALQEQKISADRYMKERSTTAVICLSVHT 91
      ||| |||:
Db      1038  GSLEQEKVVMIDLEAKRKLGGDLKLQTESIMDLNDKIQLEELKKKBE-----FDINQ 1091
      ||| |||:

Qy      92  APVEMREKLAVAEELMPRAIOELTSLNHIEEAAVLSTCNRMIEIYVVALSMNRGIREVVDW 151
      ||| |||:
Db      1092  QNSKIEDEQALALQOKKLKENQARIEELEEELEAERTAKAVEKLSRDLSELEEEISER 1151
      ||| |||:

Qy      152  MSKSG-----IPASELREHLFILRSSD--ATHPLFVSAGL-----DSLVLGE----- 193
      ||| |||:
Db      1152  LEEAGGATSVQIEMNKKREAFQKVRDRLEEAATLQHEATAALRRKHGADSVAEELGEQTDN 1211
      ||| |||:

Qy      194  -----GQILAQVKQVVRSGNSGGGLGNVI--DRMPKDAITAGKVRVS 233
      ||| |||:
Db      1212  LQRVKQKLEKESEBFKLELDDVTNMEQIIKAKANLEKVSRTLEDAQNEYVKLEEAQRS 1271
      ||| |||:

Qy      234  ETNISSGAVSVSSAAVELALMKLPKSEALSARMLLIGAGKMKL-----VIKHLVAK 285
      ||| |||:
Db      1272  LNDFTTORAKLQTEGELA-RQLEKEKALISQL-----TRGKLSYQQMEDLKRQLEEE 1324
      ||| |||:

Qy      286  GCKKVVVVNRSVERVDAIRBEMKDIIEIVRPLSDMYQAAADVVVFTSTASSETSFLF----- 341
      ||| |||:
Db      1325  GKAKNALAH-----ALQSAHRHCDLLRQEQYEEETAKAQLORVLKANSSEVAQWRTKY 1377
      ||| |||:

Qy      342  -----AKSHAEALPPVSDTMGQVR-----LFPDLSVP 368
      ||| |||:
Db      1378  ETDAIQRTEELEEAKKLQFLQDAEAAVEAVNAKCSLEKTKHRLQNEITEDLMVDVRS 1437
      ||| |||:

Qy      369  RNVSACVSEVGAARVYVVDL-----KEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDS 424
      ||| |||:
Db      1438  NAAAAALDK---KQRNFDKILAEWKQKYEESQSELESSQKEARSLSLTFKLNKVAEES 1493
      ||| |||:

Qy      425  LETVPTIKK-----LRSYADRI-----RASELEKLCQKVGEDALTQKWRRAIEELST 471
      ||| |||:
Db      1494  LEHLETFKRENKVLQBEISDLTEQLGGGKKNVHELEKVRKQLEVEKL--ELQSALESEAA 1551
      ||| |||:

Qy      472  GIVN---KLJHGPLQHLURCDGSDSRITDETLNNHALNR-----MFLSDMEKAIIIE 519
      ||| |||:
Db      1552  SLEHEECKILRAQLEFNOIKAEITERKLAEKDEEMEQAKRNHORVVDSIQSLQSLDAETSRN 1611
      ||| |||:

Qy      520  Q--KIKAKVEKTON 531

```

Db 1377 --LMVDVERSNAAAALDK---KORNFDKILAENKQYEESSQSESSQKARSLSLSTEL 1430  
QY 415 LRRFEAWRDSLETPTIKK-----LRSYADRI-----RASELEKCLQKVGEDALTKK 461  
Db 1431 FKLKWAYESLHSTFKRENKVLQEBISDLTEQLGEGGKNVHELEKIRKQLEVEKJ--E 1488  
QY 462 MRAIEELSTGIVN---KLHGPLOHLRCDGSDSRTLDTELNNMHALNR-----MF 509  
Db 1489 LQSALEAEASLEHEEGKILRAQLEFNOIKABIERKLAEKDEMEQAKRNHLRVVDSLTQ 1548  
QY 510 SLDMEKALIEQ--KIKAKVEKTON 531  
Db 1549 SLDAETRSRNEALRVKKXMEGDLN 1572

## RESULT 9

US-08-144-121-4  
; Sequence 4, Application US/08144121  
; Patent No. 5610031  
; GENERAL INFORMATION:  
; APPLICANT: Burgeson, Robert E.  
; APPLICANT: Wagnan, David W.  
; TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: BOSTON  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/144.121  
; FILING DATE: 27-OCT-1993  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Paul L.  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: (MGH-0780.0) MGP-021  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1196 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 1..250  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 251..437  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 438..807  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 808..840  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 841..1196

Query Match 4.6%; Score 119; DB 1; Length 1196;  
Best Local Similarity 19.4%; Pred. No. 0.069;

Matches 115; Conservative 87; Mismatches 209; Indels 182; Gaps 24;  
QY 28 CORVAG--GRRRS-----GVTVCDAAGVEAQA--QAVAAASVAALAEQFKI 70  
Db 513 CQCPGFGGTCSECQLFWGDPDVECRACDPRGIETPQDQSTGQCVCVGEVGPGR 572  
QY 71 SADRYMKERSITIAVIGLSVHTAPVEMREKLAVALPRAIOELTSLNH--TEEAVALST 128  
Db 573 -----DKCTRGVSGVPPDCTPCHQ-----CPALMDVIIABLTNRTHRFLEKAKALI 619  
QY 129 CNRMBIYVVALSWNRGIREVVDMMSKK-----SGIPASELRHLFIL----- 170  
Db 620 SGVIGPY-----RETVDSEVERKVSEIKDILAQSPAAPAEPLKCNIGNLPEAEKLIK 669  
QY 171 -----RSSDATRHLPEVSAGLSLVLEGQI-----LAQKQVTVSGNSQGGIG 214  
Db 670 VTEMMAQVEVKLSDTTSQSNSTAKELDSLQTEAESLQNTVKELAEQLEPKIKNSD 728  
QY 215 KNIDRMFKDAITAGKRVRSNETNISGAVSVSSA--AVELALM-----KLPKSEALSARML 267  
Db 729 DSITYFQMSLEAEERNVASTETPNSTVEQSALMRDRVEDVMMERESQFKKQESQARLL 788  
QY 268 LIGAGRMGKLV-----KHLVAKGCKKVVVNR 295  
Db 789 DELAGKLSLQLSAAAEWTCGTPGCASCSETECGGPNCRCTDEGERKCGGPGGLVTVAH 848  
QY 296 SVERVDAIREEMKDIIEIVRPLSDMYQAAADVVFTSTASETSLFAKEHAELPPVSDT 355  
Db 849 NA-----WQKAMDLD-----QDVLASALAEVQL--SKWVSEAKLRADEAKQSAE----- 890  
QY 356 MGVRLFVDISVPRNVSAVCYSEVGAARVYVDDLKEVVEANKEDRLRCAMEAQTIIITSEL 415  
Db 891 -----DILKTNAT-----KEMDKSNEELRNLIKOIRNFLTOD- 924  
QY 416 RRFEAWRDSLETPTIKKLSYADRIIRASELEKCLQKVGEDALTKMERRAIEELSTGIVN 475  
Db 925 --SADLDSIEAV-----ANEVLKMEPSTPQQL--QNLIEDIRERVESLSQ----- 966  
QY 476 KLLHGPLOHLRCDGSDSRTLDTEL--NMHALRMFSLDMKALIEQKIKAKV 526  
D, 967 --VEVILQHSAAADIARAEMLEEARAKSKSATDVKVTADVMVKEALEEAEKQV 1017

## RESULT 10

US-08-735-893-4  
; Sequence 4, Application US/08735893  
; Patent No. 5914317  
; GENERAL INFORMATION:  
; APPLICANT: Burgeson, Robert E.  
; APPLICANT: Wagnan, David W.  
; TITLE OF INVENTION: BLK CHAIN OF LAMININ AND METHODS OF USE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD  
; STREET: 60 State Street, suite 510  
; CITY: BOSTON  
; STATE: Massachusetts  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/735,893  
; FILING DATE: 18-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/144,121  
; FILING DATE: 27-OCT-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Myers, Paul L.  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: (MCH-0780.1) MGP-021DV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1196 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1..250  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 251..437  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 438..807  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 808..840  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 841..1196  
US-08-735-893-4

Query Match 4.6%; Score 119; DB 2; Length 1196;  
Best Local Similarity 19.4%; Pred. No. 0.069;  
Matches 115; Conservative 87; Mismatches 209; Indels 182; Gaps 24;  
QY 28 CORVAG-GGRRS-----GVVRCDAAGVEAQA--QAVAKAASVAALQPKI 70  
DB 513 CQMPGFGGRTCEQQLFWGPDVECRACDPRGIETPCQDQSTGQCVGVGPRC 572  
QY 71 SADRYKBERSTIAVIGLSVHTAPVEMREKLAVALAEELPRAIQELTSLNH--IEEAVALST 128  
DB 573 -----DKCTRGYGVFPDCTPCHQ-----CFALMDVITAEELNTHRFLEKAKALI 619  
QY 129 CNRMEIYVVALSWNRGIREVVDWMSK-----SGIPASLREHLFIL-----170  
DB 620 SGVIGPY-----RETVDSEKVKSEIKDILAQSPAEPPLKNIGLPEAEKLIKD 669  
QY 171 -----RSSDTRHLFEVSAGLDSLVLGEGOI-----LAQVQVVRSGQNGSLG 214  
DB 670 VTEMAQVEVKLSLDTTOSNSTAKELDSLQTEAESLDNTVKELARQLEFIKNSDIRGAL- 728  
QY 215 KNIDRMFKDAITAGKRVRSSETNISGAVSVSSA--AVEALAM-----KLPKSEALSARML 267  
DB 729 DSITYFQNSLEAEERVNASTTEPNSTVEQSALMRDRVEDVMERESQFKQEQEARLL 788  
QY 268 LIGAKMGKLV-----KHLVAKGCKKVVVYNR 295  
DB 789 DELAGKLSLDLSAAAEWTCGTPPGASCSETECGGNCKRTDBGERKCGPGCGGLVYAH 848  
QY 296 SVERVDAIREEMKDIEIVRPLSDMYQAAAEADVFTSTASETSIFAKEHAELPPVSDT 355  
DB 849 NA-----WQKAMDLD-----QDVLASALAEVEQL-SKQVSEAKLRADAEAKQSAE- 890  
QY 356 MGVRLFDVDSVPRNVACSVSGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEEL 415  
DB 891 -----DILLKTNAT-----KEKMDKNEEURNLIKQIRNFLTOD- 924  
QY 416 RRFEMRWDSLETVPITIKKLSYADIRASELEKCLQKVGEDALTKMRRRAIEELSTGIWN 475  
DB 925 ---SADLSIEAV-----ANEVLKWEPMSTPQOL--QNLTEDIRERVESLSQ----- 966  
QY 476 KLLHCPLOHLRCDSGSDRTLDTELE--NMHALNRMFSLDMEKALIEQIKAKV 526  
DB 967 ---VEVILQHSADITARAEMLEEAERAKRSKATDVKVTADMYKALEAEAKQV 1017

## RESULT 11

US-09-914-259-23  
; Sequence 23, Application US/09914259  
; Patent No. 6495336  
; GENERAL INFORMATION:  
; APPLICANT: Makowski, Lee  
; APPLICANT: Hyman, Paul  
; APPLICANT: Williams, Mark  
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
; FILE REFERENCE: 8471-010-999  
; CURRENT APPLICATION NUMBER: US/09/914.259  
; CURRENT FILING DATE: 2000-11-21  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 928  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-09-914-259-23

Query Match 4.5%; Score 118; DB 4; Length 928;  
Best Local Similarity 20.3%; Pred. No. 0.057;  
Matches 113; Conservative 95; Mismatches 200; Indels 150; Gaps 24;  
QY 7 ATTAATAATTAKPRGSSSALCQRVAGGRRRRSGVVRCDAAAGVEAQAQAAKAAVAAL 66  
DB 391 AITPSKASATTARPSTPSRLLPES-----RAETPAISDRAGTSPSLP 431  
QY 67 QPKISADRYMK-----ERSTIAVIGLSVHTAPVEMREKLAVALAEELPRAIQELTS 116  
DB 432 LKDEREEFLRENELODQIAEKESIAA-----AAERQLRE---TKSEL--IALKDHS 480  
QY 117 LNHTEAAVLGTCNRMEIYVVALSW-NRGIREVVDWMSKSGIPASELREHLFILRSS-- 173  
DB 481 KLGENERLISESENEFQMLERLAFENKEAQITIDGLKDN-----SELTAELDEVKQML 536  
QY 174 DATRHLFEVSAAGLSVLGEGOIIAQVQVVRSGQNGSLGKNIDRMFKDAITAGKRVRS 233  
DB 537 DMKMSAKETSAVLDE--KEKKAKAEKMAKMGAGDLSGDVPSDNERAVADAIA---QLDA 590  
QY 234 ETNLSGAVSVSSAAVELALMKLPSE-----ALSARMLLIGAKMGKLVKHLVA 284  
DB 591 LFEISSAGDAIPPEIDIKALREKLVETQGFVQAELSFSASASSDAEARKRAELEAR-LEA 649  
QY 285 KGCKKVVVVRNRSVERVDAIREEMKDIEIVRPLSDMYQAAAE-----ADVFTSTAS 338  
DB 650 LQOEHEELLSRNLTEAD--KEEVK--ALLAKSLSDKSQVQVLEVEQLKADIALKN--SET 703  
QY 339 SLFAKEHAELPPVSDTWGGVRLFVDISVPRNVACSVSGAARVYVNDLKEVVEANKE 398  
DB 704 -----EHLKAL-----VDDLQRRVKAGGA 722  
QY 399 DRLKAMEAQTIITEELRRFEAMRDSLETVPITIKKLSYADIRASELEKCLQKVGEDAL 458  
DB 723 G---VAMANGTVOQQLAEDFVMKSL-----MRDLQNRCEV--VELEISLDETREQ-- 770  
QY 459 TKMRRRAIEELSTGIYNKLLHGPLOHLRCDSGSDRTLDTELENMHALNRMF-----SLDM 513  
DB 771 -----YNNVLS--SNRAQQKKAFLERNLEQLTQVQRLVEQNSALKK 813  
QY 514 EKALIEQIKAKVKTQN 531  
DB 814 EVAIAERKLWARNERIQS 831

## RESULT 12

US-09-308-375-2  
; Sequence 2, Application US/09308375  
; Patent No. 6300117  
; GENERAL INFORMATION:  
; APPLICANT: Genencor International, Inc.

; TITLE OF INVENTION: Proteases From Gram-Positive Organisms  
; FILE REFERENCE: GC394-PCT  
; CURRENT APPLICATION NUMBER: US/09/308,375  
; CURRENT FILING DATE: 1999-05-14  
; EARLIER APPLICATION NUMBER: EP9719636.4  
; EARLIER FILING DATE: 1997-09-15  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 2285  
; TYPE: PRT  
; ORGANISM: Bacillus subtilis  
US-09-308-375-2

Query Match 4.5%; Score 118; DB 4; Length 2285;  
Best Local Similarity 20.7%; Pred. No. 0.24;  
Matches 107; Conservative 86; Mismatches 189; Indels 136; Gaps 21;

QY 62 VAALBQPKISADRYMKERSTIAVIGLSVHTAPVEMREKLVAEELWPRAIOELTSLNHIE 121  
DB 914 IADDDYKVAADK-----AKQSMKLAQSDIESGNKV 945  
QY 122 EAAVLSTCNREIYVVALSWNGIREVVDWMSKSGIPASELREHLFILRSSDATRHLPE 181  
DB 946 KDSVLSIANAYSSIDISNTLTSISDVVYNKMLKDDLDPELEKF-----SSSLGKLOEK 1000  
QY 182 VSAGLDSLVLGEGOILAQVQVRSQNSGGKGNIDRMFKDAITAGKVRSETNISGA 241  
DB 1001 MQKALDS---GKEAFDNAKDLQSLLETYSKSDSIDVFKMSFD-----KAQRNIDKGD 1052  
QY 242 VSVSSAAVELALMKLPKSEALSARMLLIGAGKMGKLVIKHLVAKGCKKVVVNVRSVERVD 301  
DB 1053 KSLSSV-----KSEVGDGLGETLAEGAGNEAEDFGKL-----KEALQAN-SVDDIK 1096  
QY 302 AIREMKD-----IEIYVR-----PLSDMYQAAAEADVFTSTASETSLPK 343  
DB 1097 AAIKEMSDAMQFDSVQDYLNGDIFNNTKDQVAPLNDLLEKMAEGKSI---SANEANTLIQ 1153  
QY 344 EHAELPPVSDTMGVRFLVDISVPRNVSAVSEVGAARVYVNDLKEVVE-ANK----- 397  
DB 1154 KOKELAQAISITENGUVKINRD-----EVIKORKVKLDAYNDMVYVYNNKLMKTE 1201  
QY 398 EDRLRKAMEAQTIITEELRRF-----BAWRDSLETPTIKLRSYADRIRASEL-- 446  
DB 1202 VNNAIKTLNADTLRIDSLKLRKERRKLDMSAEISDLSE---VKSINNVD--AKKELK 1255  
QY 447 --EKCLKVG-----EDALTKMRAIEELSTGIVNKLHGLPQLHRCDSG-- 490  
DB 1256 LEEKHLOPGGYSNSQIEAMQSVKSALESYISASEEATSTQEMNK-----QALVEAGTSL 1309  
QY 491 DSRITLDETLENHMLNRMFSLDMKAIIEQKIKAKVEK 528  
DB 1310 ENWTDQBEKANEETKTSYVVDKYKEALE-KVNAEIDK 1346

## RESULT 13

US-09-252-991A-18744  
; Sequence 18744, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18744  
; LENGTH: 573

; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18744

Query Match 4.5%; Score 117; DB 4; Length 573;  
Best Local Similarity 19.1%; Pred. No. 0.033;  
Matches 101; Conservative 75; Mismatches 182; Indels 170; Gaps 20;

QY 6 SATTAATAAAATYAKPRGSSSALCORVAGGRRRRSVRCDAAGVBAQAQAAKAAASVAAL 65  
DB 157 SATSARVLSLAASDP-----SAAQQLSYGSDSRQFCAMR-----EVINQLDEMEEAAAADG 208  
QY 66 EQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLVAEELWPRAIOELTSLNHIEEAAV 125  
DB 209 EASSALGERHRWQVALVAFGL-----LVLFVPGLVTRPLQRLQOR 254  
QY 126 LSTCNREIYVVALSWNGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAG 185  
DB 231 -----LVCLS-----LVLFVPGLVTRPLQRLQOR 254  
QY 186 LDSVLGEGOILAQVQVRSQNSGGKGN-----IDRMFKDAITAGKVRSETNISGA 241  
DB 255 LEEIANGDGL--RVRLEVTSRDEPGRGLSAPNAFLDKL-QPLIREVGRVGTGEVADSAGS 311  
QY 242 VSVSSAAVELALMKLPKSEALSARMLLIGAGKMGKLVIKHLVAKGCKKVVVNVRSVER-- 299  
DB 312 LAGMTAAND-----RLINSEHASVDQVSTAATOMSSAV--HEVARNAQSAQAQVADARRQA 365  
QY 300 -----VDATREEMKOTIEIVRPLSDMYQAAAEADVVFT---STASETSLPK 346  
DB 366 REGANVVEATIEVIRLAQAEVSESSIESIQLAQETASIDAVLTWIKGTAEQTNLLALNA 425  
QY 347 EALPPVSDTMGVRFLVDISVPRNVSAVSEVGAARVYVNDLKEV-----VEANKEDR 400  
DB 426 IEAARAGEQGRGFAVAD-----EVRALAARTQDSTKDIQAIERLQAGVQNA 473  
QY 401 LRKAMEAQTIITEELRRFPEARWDSLETPTIKLRSYADRIRASELEKCLKVKGEDALTK 460  
DB 474 VR-AMQSGSLKA-----RDSVE-----RAAGVGVGLAATG-DAVGR 507  
QY 461 KMRAIBELSTGIVNKLHGLPQLHRCDSGSDRTLDETLENHMLNRM 508  
DB 508 -----INDLAAQIAS-----ACE-EQSRVIDEIRNISIVREL 539

## RESULT 14

US-09-572-191-2  
; Sequence 2, Application US/09572191  
; Patent No. 6355466  
; GENERAL INFORMATION:

; APPLICANT: Beraud, Christophe  
; APPLICANT: Sakowicz, Roman  
; APPLICANT: Wood, Kenneth  
; TITLE OF INVENTION: No. 6355466el motor proteins and methods for  
; FILE REFERENCE: 1017  
; CURRENT APPLICATION NUMBER: US/09/572,191  
; CURRENT FILING DATE: 2000-05-17  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1388  
; TYPE: PRT  
; ORGANISM: Human  
US-09-572-191-2

Query Match 4.5%; Score 116.5; DB 4; Length 1388;  
Best Local Similarity 19.0%; Pred. No. 0.15;  
Matches 115; Conservative 92; Mismatches 199; Indels 199; Gaps 26;

QY 93 PVEMREKL--AAVEELWPRAIOELTSL-----NHIEAAVLSTCNREI 134  
DB 704 PPEWNEQAFPAISEEL--RTVQEQMSALQAKLDEEHNKLNKLOOHVDKLEHHSTQMGELF 761





```
Db      2 MASTTSATAGCAFAAAKTRRVGGRRPCARVAAGRRRRSGVVRCD-AGVEAQQAQA 60
Qy      56 VAKASVAALAEQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALAEELWPRAIQBEL 115
Db      61 VAKASVAALAEQFKISADRYMKERSIAVIGLSVHTAPVEMREKLAVALAEELWPRAISELT 120
Qy      116 SUNHIEEAVALSTCNRMETIYVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDA 175
Db      121 SUNHIEEAVALSTCNRMETIYVVALSWNRGLREVVDWMSKSGIPASELREHLFLMRDSDA 180
Qy      176 TRHLEFVSAGLSVLGEGQILAQVQVRSQNSGGGLGKNDIRMFKDIAITAGKVRSET 235
Db      181 TRHLEFVSAGLSVLGEGEILAQVQVRSQNSGGGLGKNDIRMFKDIAITAGKVRCE 240
Qy      236 NISSGAVSVSSAAVELALMKLPKSEALSARMILLIGAGKMGKULVIKHLVAKGCKKVVVNR 295
Db      241 NISSGAVSVSSAAVELALMKLPKSECLARMILLIGAGKMGKULVVKHLIAKAGCKKVVVNR 300
Qy      296 SVERVDATREEMKDIEIVRPLSDMYOAAAEADVFTSTASSETSLPAKEHAELPPVSDT 355
Db      301 SVERVDATREEMKDIEIVRPLSDMYOAAAEADVFTSTASSETPLTKEHAELPAISDA 360
Qy      356 MGVRLFVDISVPRNVSCVSEVGARVYVVDLKEVVEANKEDRLRKAMEAQTIITBEL 415
Db      361 MGVRLFVDISVPRNVSCVSEVGARVYVVDLKEVVEANKEDRLRKAMEAQTIITBEL 420
Qy      416 RRFEARWRSLEVTPIKRLRSVADRIRASELEKCLQKVGEDALTKKMRRAIEELSTGIVN 475
Db      421 RRFEARWRSLEVTPIKRLRSVADRIRASELEKCLQKVGEDALTKKTRRSIEELSTGIVN 480
Qy      476 KLLHGPLOHLRCGSDSRITLDTLENMHALNMFSLDMEKATIEOKIKAKVEKTON 531
Db      481 KLLHGPLOHLRCGSDSRITLDTLENMHALNMFSLDMEKATIEOKIKAKVEKSON 536

RESULT 2
ID      HMA1_HORVU STANDARD; PRT; 527 AA.
AC      Q42843; Q42844;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Glutamyl-tRNA reductase 1, chloroplast precursor (BC 1.2.1.-) (Glutr).
OS      HMA1.
GN      Hordeum vulgare (Barley).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC      Triticeae; Hordeum.
OX      NCBI_TaxID=4513;
RN      [1]
RN      [1] SEQUENCE FROM N.A.
RC      STRAIN=cv. Klages;
RX      MEDLINE=96267576; PubMed=8696365;
RA      Bougri O., Grimm B.;
RT      "Members of a low-copy number gene family encoding glutamyl-tRNA
RT      reductase are differentially expressed in barley."
RL      Plant J. 9:867-878(1996).
RN      [2]
RN      [2] SEQUENCE OF 44-65, AND CHARACTERIZATION.
RX      MEDLINE=95045501; PubMed=7957167;
RA      Pontoppidan B., Kannangara C.G.;
RT      "Purification and partial characterisation of barley glutamyl-
RT      tRNA(Glu) reductase, the enzyme that directs glutamate to chlorophyll
RT      biosynthesis."
RL      Eur. J. Biochem. 225:529-537(1994).
RN      [3]
RN      [3] 3D-STRUCTURE MODELING OF 76-460.
RX      MEDLINE=20058652; PubMed=10591107;
RA      Brody S.S., Gough S.P., Kannangara C.G.;
RT      "Predicted structure and fold recognition for the glutamyl tRNA
RT      reductase family of proteins."
RL      Proteins 37:485-493(1999).
CC      -!- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
semialdehyde + NADP(+) + tRNA(Glu).
-!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
Involved in chlorophyll biosynthesis.
-!- SUBCELLULAR LOCATION: Chloroplast.
-!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X92403; CAA63140.1; -
EMBL; X86101; CAA60054.1; ALT_INIT.
PDB; 1B29; 09-DEC-98.
PDB; 1B61; 18-FEB-00.
InterPro; IPR000343; Glutr.
Pfam; PF00745; Glutr_dimer; 1.
Pfam; PF05201; Glutr_N; 1.
Pfam; PF05200; Glutr_NAD_bind; 1.
TIGRFAMs; TIGR01035; hemA; 1.
PROSITE; PS00747; Glutr; 1.
Porphyryn biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
NADP; Transit peptide; Chloroplast; Multigene family; 3D-structure.
CHAIN 44 527
TRANSIT 1 43
DOMAIN 286 289
STRAND 79 91
STRAND 96 99
HELIX 104 125
TURN 126 126
STRAND 133 136
TURN 139 140
TURN 150 152
TURN 155 159
TURN 160 160
TURN 170 182
TURN 192 192
TURN 193 197
TURN 205 206
TURN 209 213
TURN 225 230
TURN 233 251
TURN 252 253
TURN 259 260
TURN 262 265
TURN 269 281
TURN 282 282
TURN 285 288
TURN 293 302
TURN 303 303
TURN 305 306
TURN 309 309
TURN 313 315
TURN 316 318
TURN 319 321
TURN 325 327
TURN 339 340
TURN 345 345
TURN 347 350
TURN 351 352
TURN 357 358
TURN 365 365
TURN 368 376
TURN 377 378
TURN 380 381
TURN 384 385
TURN 386 389
TURN 390 390
TURN 398 404
TURN 405 405
```

FT HELIX 413 421  
 FT TURN 422 422  
 FT HELIX 423 436  
 FT TURN 437 438  
 FT HELIX 443 457  
 FT TURN 458 459  
 SQ SEQUENCE 527 AA; 57652 MW; D329E56459BE1165 CRC64;  
 Query Match 88.8%; Score 2313; DB 1; Length 527;  
 Best Local Similarity 88.1%; Pred. No. 2.1e-119; Indels 4; Gaps 2;  
 Matches 468; Conservative 27; Mismatches 32; Indels 4; Gaps 2;  
 QY 1 MATTTTAAATAAATTAAPRSGSSALCORVAGGRRRSGVVRCDAGG---DAQAASRAA 60  
 DB 1 MAGATSA-TAAAGAFAPAAKARGPAAACPLWVAAGRRRSGVVRCDAGG---DAQAASRAA 56  
 QY 61 SVAALAEOKISADRYMKEKSTIAVIGLSVHTAPVEMREKLAVAEELWPAQELTSLNHI 120  
 DB 57 SITALEOQKISADRYMKEKSSIAVIGLSVHTAPVEMREKLAVAEELWPAQELTSLNHI 116  
 QY 121 EEAVALSTCNRMIEIYVVALSWNRGIREVVDMWSKSGIPASELREHLFILRSSDATRHLF 180  
 DB 117 EEAVALSTCNRMIEIYVVALSWNRGIREVVDMWSKSGIPASELREHLFILRSSDATRHLF 176  
 QY 181 EVSAGLSLVLGEGQILAQVKQVVRSGNSGGLGNIDRMFKDAITAGKRVRSSETNISG 240  
 DB 177 EVSAGLSLVLGEGQILAQVKQVVRSGNSGGLGNIDRMFKDAITAGKRVRSSETNISG 236  
 QY 241 AVSVSSAAVELALMKLPKSEALSARMLLIGAGMGKLVIKHLVAGCKKVVVVRNRSVRV 300  
 DB 237 AVSVSSAAVELALMKLPKSEALSARMLLIGAGMGKLVIKHLVAGCKKVVVVRNRSVRV 296  
 QY 301 DAIREEMKDIEIVRPLSDMYQAAAEADVFTSTASSETSLFAKEHAEALPPVSDTMGVVR 360  
 DB 297 DAIREEMKDIEIVRPLSDMYQAAAEADVFTSTASSETSLFAKEHAEALPPVSDTMGVVR 356  
 QY 361 LFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEELRRPEA 420  
 DB 357 LFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEELRRPEA 416  
 QY 421 WRDSLETPTTIKRLSYADRIASELEKLCQKVGEDALTQKRRRAIELSTGIYVKNLHG 480  
 DB 417 WRDSLETPTTIKRLSYADRIASELEKLCQKVGEDALTQKRRRAIELSTGIYVKNLHG 476  
 QY 481 PLOHLRCDGSDSRITDELLENHNLNRMFSLDMKAEIIEQIKKAKVEKTON 531  
 DB 477 PLOHLRCDGSDSRITDELLENHNLNRMFSLDMKAEIIEQIKKAKVEKTON 527

## RESULT 3

ID HMA3 HORVU STANDARD; PRT; 535 AA.  
 AC O65796;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamyl-tRNA reductase 3, chloroplast precursor (EC 1.2.1.-) (Glutr).  
 GN HMA3.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
 OC Triticeae; Hordeum.  
 OK NCBI\_TaxID=4513;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=cv. Bonus; TISSUE=Root;  
 RA Tanaka R., Yoshida K., Nakayaahiki T., Tsuji H., Inokuchi H.,  
 RA Okada K., Tanaka A.;  
 RA "The third member of the hema gene family encoding glutamyl-tRNA  
 RT reductase is primarily expressed in roots in Hordeum vulgare."  
 RL Photosyn. Res. 53:161-171(1997).  
 CC -1- CATALYTIC ACTIVITY: Glutamyl-tRNA (Glu) + NADPH = glutamate-1-  
 semialdehyde + NADP(+) + tRNA (Glu).

CC -1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.  
 CC Involved in chlorophyll biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: Chloroplast.  
 CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN ROOTS.  
 CC -1- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; D88383; BAA25168.1; -;  
 CC FIR; T04402; T04402.  
 CC HSSP; Q42843; 1B29.  
 CC InterPro; IPR000343; Glutr.  
 CC InterPro; IPR000205; NAD binding.  
 CC InterPro; IPR000594; ThiF domain.  
 CC Pfam; PF00745; Glutr\_dimer; 1.  
 CC Pfam; PF05201; Glutr\_N; 1.  
 CC Pfam; PF05200; Glutr\_NAD\_bind; 1.  
 CC TIGRFAMs; TIGR01035; hema; 1.  
 CC PROSITE; PS00747; Glutr; 1.  
 CC Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;  
 CC NADP; Transit peptide; Chloroplast; Multigene family.  
 FT TRANSIT 1 CHLOROPLAST (POTENTIAL).  
 FT CHAIN ? 535 GLUTAMYL-TRNA REDUCTASE 3.  
 SQ SEQUENCE 535 AA; 58419 MW; B2830889AE6A3224 CRC64;  
 Query Match 85.4%; Score 2225; DB 1; Length 535;  
 Best Local Similarity 84.0%; Pred. No. 1.4e-114;  
 Matches 452; Conservative 38; Mismatches 38; Indels 10; Gaps 5;  
 QY 1 MATTTTAAATAAATAA---KPRGSSALCORVAGGRRR-SGVVRCDAGVQAQA 55  
 DB 1 MASTSTASATAGAFAPAAAGVYKPRG--SAACPRVPAGRGRLSCVVRCD-AGPGVPAQM 57  
 QY 56 VAKAASVAALAEOKISADRYMKEKSTIAVIGLSVHTAPVEMREKLAVAEELWPAQELT 115  
 DB 58 AAMAASVAALAEOKISADRYMKEKSSIAVIGLSVHTAPVEMREKLAVAEELWPAQELT 117  
 QY 116 SLNHIEAVALSTCNRMIEIYVVALSWNRGIREVVDMWSKSGIPASELREHLFILRSSDA 175  
 DB 118 NLNHIEAVALSTCNRMIEIYVVALSWNRGIREVVDMWSKSGIPASELREHLFILRSSDA 177  
 QY 176 TRHLFEVSAGLSLVLGEGQILAQVKQVVRSGNSGGLGNIDRMFKDAITAGKRVRSSET 235  
 DB 178 TRHLFEVSAGLSLVLGEGQILAQVKQVVRSGNSGGLGNIDRMFKDAITAGKRVRSSET 237  
 QY 236 NISSGAVSVSSAAVELALMKLPKSEALSARMLLIGAGMGKLVIKHLVAGCKKVVVYNR 295  
 DB 238 NISCGAVSVSSAAVELALMKLPKSEALSARMLLIGAGMGKLVAKHLAAKCKKVVYNR 297  
 QY 296 SVERVDAIREMKDIEIVRPLSDMYQAAAEADVFTSTASSETSLFAKEHAEALPPVSDT 355  
 DB 298 SVERVDAIREMKDIEIVRPLSDMYQAAAEADVFTSTASSETSLFAKEHAEALPPVSDT 357  
 QY 356 MGVVRLFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTIITEEL 415  
 DB 358 LGGVRLFVDISVPRNVSACVSDVGHARVYVNDLKEVVEANKEDRLRKAMEAQTIITEEL 417  
 QY 416 RRFEMARDSLETPTTIKRLSYADRIASELEKLCQKVGSDALTQKRRRAIELSTGIYN 475  
 DB 418 KRFEAMRDSMETVPTTIKRLSYADRVRASELKCLQKIGEDALTQKRRRSIEQLSTGIYN 477  
 QY 476 KLLHGPLOHLRCDGSDSRITDELLENHNLNRMFSLDMKAEIIEQIKKAKVE--KTON 531  
 DB 478 RLLHGPLOHLRCDGSDSRITDELLENHNLNRMFGLDTEKAVMEQIKTKVKEKQKTON 535

## RESULT 4



HMA2\_HORVU STANDARD; PRT; 465 AA.

ID HMA2\_HORVU AC Q96563; DT 01-NOV-1997 (Rel. 35, Created) DT 01-NOV-1997 (Rel. 35, Last sequence update) DT 28-FEB-2003 (Rel. 41, Last annotation update) DE Glutamyl-tRNA reductase 2 (EC 1.2.1.-) (Glutr) (Fragment). GN HMA2.

OS Hordeum vulgare (Barley). OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; OC Triticeae; Hordeum. OX NCBI\_TaxID=4513; RN [1] SEQUENCE FROM N.A. RP STRAIN=cv. Bonus; RX MEDLINE=96267576; PubMed=8696365; RA Bougri O., Grimm B.; RT "Members of a low-copy number gene family encoding glutamyl-tRNA reductase are differentially expressed in barley." RL Plant J. 9:867-878(1996). CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA (Glu) + NADPH = glutamate-1-semialdehyde + NADP(+) + tRNA(Glu). CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step. CC Involved in chlorophyll biosynthesis. CC -!- SUBCELLULAR LOCATION: Chloroplast. CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X86102; CAA60055.1; . DR PIR; T05734; T05734. DR HSSP; Q42843; 1B29. DR InterPro; IPR000343; Glutr. DR Pfam; PF00745; Glutr dimer; 1. DR Pfam; PF05201; Glutr N; 1. DR Pfam; PF05200; Glutr NAD bind; 1. DR TIGRfams; TIGR01035; hema; 1. DR PROSITE; PS00747; Glutr; 1. DR Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase; KW NADP; Chloroplast; Multigene family. FT NON TER 1 1 POLY-VAL. FT DOMAIN 224 227 FT SEQUENCE 465 AA; 51674 MW; E42A50A3E749CE6D CRC64;

Query Match  
Best Local Similarity 81.1%; Score 2113; DB 1; Length 465;  
Matches 418; Conservative 29; Mismatches 18; Indels 0; Gaps 0;

QY 67 QFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALPRAITOLTSLNHIEAAVL 126  
DB 1 QFKISADRYIEKSSIAVIGLSVHTAPVDMREKLAVALPRAISELTSLNHIEAAVL 60

QY 127 STCNRMETIYVVALSWNRGIREVDMWMSKSGIPASELREHLFILRSSDATHLFEVSAGL 186  
DB 61 STCNRMETIYVVALSWNRGIREVDMWMSKSGIPASELREHLFILRSSDATHLFEVSAGL 120

QY 187 DSLVLGEGQIIAQQVQVRSNGSGGLGKNIDRMFKDAITAGKRVSRNTSSGAVSVSS 246  
DB 121 DSLVLGEGQIIAQQVQVRSNGSGGLGKNIDRMFKDAITAGKRVSRNTSSGAVSVSS 180

QY 247 AAVELALMKLPKSEALSARMILLIGAKMGKLVKHLVAKGCKVWVNRVSDATREE 306  
DB 181 AAVELALMKLPKSEALSARMILLIGAKMGKLVKHLVAKGCKVWVNRVSDATREE 240

QY 307 MKQIEIVRPLSDMYQAAAEADVFTSTASTSIFAKEHAALPVPVSDTMGGVRLFVDIS 366  
DB 1 MKQIEIVRPLSDMYQAAAEADVFTSTASTSIFAKEHAALPVPVSDTMGGVRLFVDIS 366

DB 241 MKQIEIVRPLSDMYQAAAEADVFTSTASTSIFAKEHAALPVPVSDTMGGVRLFVDIS 300

QY 367 VPRNVSAVSEVGAARVYVDDLKSEVVEANKEDRLKAMEAQTIITEELRFEAWRDSLE 426  
DB 301 VPRNVGACLSQVEHARVYVDDLKSEVVEANKEDRVKAMEAQAIITQELKRFEAWRDSLE 360

QY 427 TVPTIKLRSYADIRASELEKCLQKVGEDALTCKMRAIBELSTGIVNKLHGPLEHLR 486  
DB 361 TVPTIKLRSYADIRASELEKCLQKVGEDALTCKMRAIBELSTGIVNKLHGPLEHLR 420

QY 487 CDGSDSRITLDTLENMHALNMFSLDMKAIIEOKIKAKVKTKTON 531  
DB 421 CDGSDSRITLDTLENMHALNMFSLDMKAIIEOKIKAKVKTKTON 465

RESULT 5  
HMA1\_CUGSA STANDARD; PRT; 552 AA.

ID HMA1\_CUGSA AC P93111; DT 01-NOV-1997 (Rel. 35, Created) DT 01-NOV-1997 (Rel. 35, Last sequence update) DT 28-FEB-2003 (Rel. 41, Last annotation update) DE Glutamyl-tRNA reductase 1, chloroplast precursor (EC 1.2.1.-) (Glutr). GN HMA1. OS Cucumis sativus (Cucumber). OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC eurosida I; Cucurbitales; Cucurbitaceae; Cucumis. OX NCBI\_TaxID=3659; RN [1] SEQUENCE FROM N.A. RP STRAIN=cv. Aonagajibai; TISSUE=Cotyledon; RX MEDLINE=97088717; PubMed=8934625; RA Tanaka R., Yoshida K., Nakayashiki T., Masuda T., Tsuji H., RA Inokuchi H., Tanaka A.; RT "Differential expression of two hema mRNAs encoding glutamyl-tRNA reductase proteins in greening cucumber seedlings." RL Plant Physiol. 110:1223-1230(1996). CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA (Glu) + NADPH = glutamate-1-semialdehyde + NADP(+) + tRNA(Glu). CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step. CC Involved in chlorophyll biosynthesis. CC -!- SUBCELLULAR LOCATION: Chloroplast (By similarity). CC -!- TISSUE SPECIFICITY: PRIMARILY IN COTYLEDONS AND HYPOCOTYLS OF GREENING CUCUMBER SEEDLINGS. CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D50407; BAA08910.1; . DR PIR; T10186; T10186. DR HSSP; Q42843; 1B29. DR InterPro; IPR000343; Glutr. DR Pfam; PF00745; Glutr dimer; 1. DR Pfam; PF05201; Glutr N; 1. DR Pfam; PF05200; Glutr NAD bind; 1. DR TIGRfams; TIGR01035; hema; 1. DR PROSITE; PS00747; Glutr; 1. DR Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase; KW NADP; Transit peptide; Chloroplast; Multigene family. FT TRANSIT 1 2 CHLOROPLAST (POTENTIAL). FT CHAIN ? 552 GLUTAMYL-TRNA REDUCTASE 1. FT SEQUENCE 552 AA; 60895 MW; 9C7A72AF24FF00F5 CRC64;

Query Match  
Best Local Similarity 68.5%; Score 1784.5; DB 1; Length 552;  
Matches 367; Conservative 74; Mismatches 86; Indels 27; Gaps 8;

QY 1 MATTTSATTA-----AAAATTAKPRGSSA-----LCORVAGGRR-----BSGV----- 41  
 DB 1 MAVSTFSQAKUEALLFKASNSSTNLSHLPFCFKSIR--TRRILFQRTGVSSFTP 58  
 QY 42 VRCDAAGVEAQAQ-----AVAKAASVAALBQFKISA-DRYMKERSTIAVIGLSVHTAPVEM 96  
 DB 59 PKCELASSDVLVQNDIEDPPKSNLSALBOLKTSADVRYTKERSIIVIGLSIHTTPVEM 118  
 QY 97 REKLVAELWPRATOELTSLNHIREEAVALSTCNMEIYVVALSNWRGIREVVDWMSKS 156  
 DB 119 REKLVAELWPRATOELTSLNHIREEAVALSTCNMEIYVVALSNWRGIREVVDWMSKS 178  
 QY 157 GIPASELREHLFILRSSDATRHLFEVSAGLSILVLGEGQILAQVKQVVRSGNSGGLGN 216  
 DB 179 GIPVSEICQHRELLYNNDATQHI FEVSAGLSILVLGEGQILAQVKQVVRSGNSGGLGN 238  
 QY 217 IDRMFKDAITAGKVRSETNISGAVSVSSAAVELALMKLPKSEALSARMLLIGAGKMK 276  
 DB 239 ISGLFKHAIIVGKRVRTETNIAAGAVSVSSAAVELALMKLPKSEALSARMLLIGAGKMK 298  
 QY 277 LVKHLVAGCKKVVVNRVSVERVDAIREEMKDIEIVRPLSDMYQAAAEADVVTSTAS 336  
 DB 299 LVKHLVAGCKKVVVNRVSVERVDAIREEMKDIEIVRPLSDMYQAAAEADVVTSTAS 358  
 QY 337 ETSLFAKHAELPVPVDTMGVRLFVDISVPRNVSACVSEYGAARVYNVDLKEVZAN 396  
 DB 359 ESSLFTKEQKDLPPVGHVGVGLRFLFIDISVPRNVSACVSEYGAARVYNVDLKEVZAN 418  
 QY 397 KEDRLRKAMEATITTELRPEARWRDSLETPTTKLRSYADRIRASELEKCLKOVGED 456  
 DB 419 KEDRLRKAMEATITTELRPEARWRDSLETPTTKLRSYADRIRASELEKCLKOVGED 478  
 QY 457 ALTKKWRRAIELSTGVNKLHGLPDLHLCRDSRSLDETLNHNHNRNFSLDMEKA 516  
 DB 479 -IPKTRRAVDDLSGIVNKLHGLPDLHLCRDSRSLDETLNHNHNRNFSLDMEKA 537  
 QY 517 IIEQKIKAKVEKTK 530  
 DB 538 VLEQKIRAKVEQNK 551

## RESULT 6

ID HMAI ARATH STANDARD; PRT; 543 AA.  
 AC P42804; Q9LOB9;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamyl-tRNA reductase 1, chloroplast precursor (EC 1.2.1.1-) (Glutr).  
 GN HEMA1 OR HEMA OR ATIG58290 OR F19C14.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE=Leaf;  
 EX MEDLINE=94198598; PubMed=7808550;  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,

RA Gill J.B., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.B.,  
 RA Kim C.J., Koo H.B., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,  
 RA Milgitscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:816-820(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the  
 RT SSP consortium (Salk/Stanford/PGEC).";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 CC -|- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-  
 CC semialdehyde + NADP(+) + tRNA(Glu).  
 CC -|- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.  
 CC Involved in chlorophyll biosynthesis.  
 CC -|- SUBCELLULAR LOCATION: Chloroplast.  
 CC -|- TISSUE SPECIFICITY: PRESENT IN ALL TISSUES TESTED. SLIGHTLY MORE  
 CC ABUNDANT IN LEAVES.  
 CC -|- INDUCTION: By light.  
 CC -|- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U03774; AAA19118.1; -;  
 CC EMBL; AC008051; AAF02258.1; -;  
 CC EMBL; AY072223; AAL60044.1; -;  
 CC PIR; E96616; E96616.  
 CC HSSP; Q42843; 1B29.  
 CC InterPro; IPR000343; Glutr.  
 CC Pfam; PF00745; Glutr dimer; 1.  
 CC Pfam; PF05201; Glutr N; 1.  
 CC Pfam; PF05200; Glutr NAD bind; 1.  
 CC TIGRFAMs; TIGR01035; hemA; 1.  
 CC PROSITE; PS00747; GLUTR; 1.  
 CC Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;  
 CC NADP; Transit peptide; Chloroplast; Multigene family.  
 CC CHLOROPLAST (POTENTIAL).  
 CC GLUTAMYL-TRNA REDUCTASE 1.  
 CC CHAIN ? 543  
 CC DOMAIN 23 33  
 CC POLY-SER.  
 CC POLY-VAL.  
 CC E -> V (IN REF. 1).  
 CC K -> N (IN REF. 1).  
 CC K -> L (IN REF. 1).  
 CC Q -> H (IN REF. 1).  
 CC CONFLICT 169  
 CC CONFLICT 291  
 CC CONFLICT 542  
 CC SEQUENCE 543 AA; 59515 MW; 04A095FEC96CC014 CRC64;  
 Query Match 67.9%; Score 1770.5; DB 1; Length 543;  
 Best Local Similarity 69.5%; Pred. No. 9e-90;  
 Matches 360; Conservative 59; Mismatches 86; Indels 13; Gaps 4;  
 QY 23 SSSALCORVAG-----GGRERSGVVRCDAAGVRAQAQAAKAAVALEQFKIS-A 72  
 DB 28 SSSSVSGTPLGNGVRVLPKNRTRRGLIQKARCELSASSDSNAASIALEQKNSAA 87  
 QY 73 DRYMKERSTIAVIGLSVHTAPVREKLAVALAEELWPRAIQELTSLNHIREEAVALSTCNRM 132  
 DB 88 DRYTKERSIIVIGLSIHTAPVREKLAVALAEELWPRAIQELTSLNHIREEAVALSTCNRM 147

```

QY 133 EIIYVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLSVLG 192
DB 148 EIIYVALSQRHGRVKEVTEWMSKTSIPVSEICQHRFLYNDKATQHI FEVSAGLSVLG 207
QY 193 EQGIIAQVQVVRSGNSGLCKNTDRMPKDAITAGKRVSRSTNITSSGAVSVSSAAVELA 252
DB 208 EQGIIAQVQVVRSGNSGLCKNTDRMPKDAITAGKRVSRSTNITSSGAVSVSSAAVELA 267
QY 253 LMKLPKSEALSARMILLIGAKMGKLVIKHLVAKGCKKVVVNRSVRVDAREEMKDIEI 312
DB 268 LMKLPQSSNVSAARMCVIGAKMGKLVIKHLVAKGCKKVVVNRSEERSVSAIREEMPGIEI 327
QY 313 VYRPLSDMYQAAAEADVFTSTASETSLFAKEHAALPVSTMGVRLFVDISVPRNVS 372
DB 328 IYRPLDEMLACASEADVFTSTASETSLFAKEHAALPVSTMGVRLFVDISVPRNVS 387
QY 373 ACVSEVGARVYVVDLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETVPYTIK 432
DB 388 SCVGEVETARVYVVDLKEVVAANKEDRMKAMEAQTIITEESTQFEAWRDSLETVPYTIK 447
QY 433 KLRVSADIRASELEKCLKQVGEDALTUKORRAIBELSTGIVNKLHGLPQLHRCGSDS 492
DB 448 KLRVAERIRVAELEKCKSKMGDD-INKKTTRAVDDLSRGIVNRLHGLPQLHRCGSDS 506
QY 493 RTLDETLENHNLNMFSLDMEKALIEQKIKAKVKTQ 530
DB 507 RTLSETLENHNLNRMIGL--EKDILEKIKAMAEQQQ 542

RESULT 7
HMA2 ARATH
ID HMA2 ARATH STANDARD; PRT; 530 AA.
AC P49294; O04950;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamyl-tRNA reductase 2, chloroplast precursor (EC 1.2.1.-) (Glutr).
GN HMA2 OR AT1G09940 OR F21M12.33.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
ON NCBI_TaxID=3702;
RX SEQUENCE FROM N.A.
RC STRAIN=sv. Columbia;
RC MEDLINE=96189258; PubMed=8605295;
RA Kumar A.M., Ceankovszki G., Soell D.;
RT "A second and differentially expressed glutamyl-tRNA reductase gene
RL from Arabidopsis thaliana."
RN Plant Mol. Biol. 30:419-426(1996).

SEQUENCE FROM N.A.
RC STRAIN=sv. Columbia;
RC MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egtu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizier L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

```

```

thaliana.";
RL Nature 408:816-820(2000).
CC -! CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA(Glu).
CC -! PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC -! Involved in chlorophyll biosynthesis.
CC -! SUBCELLULAR LOCATION: Chloroplast (By similarity).
CC -! TISSUE SPECIFICITY: EXPRESSED IN LOW LEVELS IN ROOTS AND FLOWERS.
CC -! SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U27118; AAB01674.1; -.
CC EMBL; AC000132; AAB00749.1; -.
CC PIR; G86233; G86233.
CC PIR; S65773; S65773.
CC HSSP; Q42843; 1B29.
CC InterPro; IPR000343; Glutr.
CC InterPro; IPR000594; Thif domain.
CC Pfam; PF00745; Glutr_dimer; 1.
CC Pfam; PF05201; Glutr_N; 1.
CC Pfam; PF05200; Glutr_NAD_bind; 1.
CC TIGRFAMs; TIGR01035; hemA; 1.
CC PROSITE; PS00747; GLUTR; 1.
CC Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
KW NADP; Transit peptide; Chloroplast; Multigene family.
FT TRANSIT 1 64 CHLOROPLAST (POTENTIAL).
FT CHAIN 65 530 GLUTAMYL-TRNA REDUCTASE 2.
FT CONFLICT 410 410 E -> D (IN REF. 2).
SQ SEQUENCE 530 AA; 58305 MW; 0CB225296BA3D1A8 CRC64;

Query Match 64.5%; Score 1680; DB 1; Length 530;
Best Local Similarity 67.2%; Pred. No. 7.6e-85;
Matches 350; Conservative 60; Mismatches 85; Indels 26; Gaps 9;

QY 20 PRGSSALCQVAG-----GRRRSVGV---RCDAGVEAQAAVAAVAALEQFKI 70
DB 23 PTYSSSPAPLDVIGIRALPMNRKRGIIQARCEISPSN-----KAASIALEQKT 75
QY 71 SA-DRYMKERSTIAVIGLSVHTAPVEMREKLAVALAEELMPRAIQELTSLNHIEAAVLSTC 129
DB 76 SAIDRYTKERSIIVVIGLSIHTAPVEMREKLAIPAEWPRALAEICGLNHIEAAVLSTC 135
QY 130 NRMEIYVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLSL 189
DB 136 NRMEIYVVALSQRHGRVKEVTEWMSKTSIPVSEICQHRFLYNDKATQHI FEVSAGLSL 195
QY 190 VLGEQIIAQVQVVRSGNSGLCKNTDRMPKDAITAGKRVSRSTNITSSGAVSVSSAAV 249
DB 196 VLGEQIIAQVQVVRSGNSGLCKNTDRMPKDAITAGKRVSRSTNITSSGAVSVSSAAV 255
QY 250 ELALMKLPK-SEALSARMILLIGAKMGKLVIKHLVAKGCKKVVVNRSVRVDAREEM- 307
DB 256 ELALMKLPKSEASHSARMVLVVGAGMGKLVIKHLVAKGCTQWVVVNRSEKVAARVEMP 315
QY 308 KDIEIYRPLSDMYQAAAEADVFTSTASETSLFAKEHAALPVSTMGVRLFVDISV 367
DB 316 PGVEIITKPLDEMLSCAAEADVFTSTASETSLFAKEHAALPVSTMGVRLFVDISV 371
QY 368 PRNVSACVSEVGARVYVVDLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLET 427
DB 372 PRNVSACVSEVGARVYVVDLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLET 431
QY 428 VPTIKLRSYADIRASELEKCLKQVGEDALTUKORRAIBELSTGIVNKLHGLPQLHRC 487
DB 432 VPTIKLRSYADIRASELEKCLKQVGEDALTUKORRAIBELSTGIVNKLHGLPQLHRC 490

```



```

DR InterPro: IPR000343; GlutR.
DR Pfam: PF00745; GlutR dimer; 1.
DR Pfam: PF05201; GlutR_N; 1.
DR Pfam: PF05200; GlutR_NAD bind; 1.
DR TIGRFAMs: TIGR01035; hemaA; 1.
DR PROSITE: PS00747; GlutR; 1.
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
FT ACT_SITE 50 50
FT ACT_SITE 99 99
FT ACT_SITE 99 99
FT CONFLICT 33 33
FT CONFLICT 33 33
FT CONFLICT 48 48
FT CONFLICT 174 174
FT CONFLICT 280 280
FT CONFLICT 374 380
FT CONFLICT 374 380
SQ SEQUENCE 428 AA; 47909 MW; 06A470955F9D3C64 CRC64;

Query Match 34.9%; Score 909.5; DB 1; Length 428;
Best Local Similarity 47.0%; Pred. No. 6.6e-43;
Matches 204; Conservative 77; Mismatches 138; Indels 15; Gaps 7;

QY 82 IAVIGLSVHTAPVEMREKLAVAEELWPAIQLTSLNHIEEAVALSTCNRMIEYVALSW 141
DB 3 IAVVGLSHKTAPEVEIREKLSIQEPQTESAIAQLTSYPHIDEVAILSTCNRLIYIVAGET 62

QY 142 NRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQVK 201
DB 63 DHGIREVTOFUSEHSLPWSLURQLFVLLHEDAVHMRVAAGLDSLVLGEGQILAQVK 122

QY 202 QVVRSGQSGGLGKNDIRMFKDITAGKRVSRSETNISSGAVSVSSAAVELALMKLPKSEA 261
DB 123 NTHKLGQQYNGIKTILNRLFRQALTAGKRVRTETSIGTGVNISAAVELAQI---KAEN 179

QY 262 LSA-RMLLIGAGKMKLVKHLVAGCKKVVVNRVSRVDAIREMKDIEIVRPLSDM 320
DB 180 LAACRVTLGAGKMSRLVQHLVSKGATQISIVNRSRERAQELAKQFSEHPIRTHLPEM 239

QY 321 YQAAEADVFTSTASSETSLFAKEHAE-ALPPVSDTMGGVRLFDVDSVPRNVSAACVSEVG 379
DB 240 MVTVAESHLVFTSTASPELDRLAKLEMLAPNQPLM---LF-DISVPRNVHTDVNBLV 294

QY 380 AARVYNVDLKEVBEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETVTPIKKLSYAD 439
DB 295 NVQAFNVDDLKAVVAONYESRRKMAQEAERLEIEAFDIWRSLETVSTISLSRSKIE 354

QY 440 RIRASELCKLOKVEDALTKWRRAIEBELSTGIYKVLHGLHQLRCOGSDSRITLDEL 499
DB 355 TIREOELEKALSRLGSE-FGDKHQEVEIATRGIVNKLHDPVQLRA---QQDVEARR 409

QY 500 ENMHALNRMFSLDM 513
DB 410 RCMQTLQWLFNLVDV 423

RESULT 10
HEMI_SYNY3
ID -HEMI_SYNY3 STANDARD; PRT; 427 AA.
AC P28463;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GlutR).
GN HEMA OR SLR1808.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxId=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92235044; PubMed=1569081;
RA Verkamp E., Jahn M., Jahn D., Kumar A.M., Soell D.;
RT "Glutamyl-tRNA reductase from Escherichia coli and Synecocystis
RT 6803. Gene structure and expression.";
RL J. Biol. Chem. 267:8275-8280(1992).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93093989; PubMed=1459859;
RA Grimm B.;
RT "Identification of a hema gene from Synecocystis by complementation
RT of an E. coli hema mutant";
RL Hereditas 117:195-197(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuura A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA(Glu).
CC -1- ENZYME REGULATION: FEEDBACK INHIBITION BY HEME.
CC -1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC Involved in chlorophyll biosynthesis.
CC -1- SUBUNIT: Homooctamer.
CC -1- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84218; AAA27289.1; -.
DR EMBL; X65963; CAA46779.1; -.
DR EMBL; D90908; BAA17738.1; ALT_INIT.
DR HSP; Q42843; 1B29.
DR HAWAP; MF_00087; -.
DR InterPro: IPR000343; GlutR.
DR InterPro: IPR000594; Thif domain.
DR Pfam: PF00745; GlutR_dimer; 1.
DR Pfam: PF05201; GlutR_N; 1.
DR Pfam: PF05200; GlutR_NAD bind; 1.
DR TIGRFAMs; TIGR01035; hemaA; 1.
DR PROSITE; PS00747; GlutR; 1.
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
FT ACT_SITE 50 50
FT ACT_SITE 99 99
FT ACT_SITE 99 99
SQ SEQUENCE 427 AA; 47525 MW; 5F5D9D744A900D7D CRC64;

Query Match 34.7%; Score 904; DB 1; Length 427;
Best Local Similarity 44.7%; Pred. No. 1.3e-42;
Matches 194; Conservative 81; Mismatches 147; Indels 12; Gaps 4;

QY 82 IAVIGLSVHTAPVEMREKLAVAEELWPAIQLTSLNHIEEAVALSTCNRMIEYVALSW 141
DB 3 IAVVGLSHKTAPEVEIREKLSIQEALREALTSLKSYPHIEBVTIVISCNRLIYAVVTD 62

QY 142 NRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQVK 201
DB 63 EKVVEITQFLSETGNIPLATLRRYLFTLLHEDAVHMRVAAGLDSLVLGEGQILAQVR 122

QY 202 QVVRSGQSGGLGKNDIRMFKDITAGKRVSRSETNISSGAVSVSSAAVELALMKLPKSEA 261
DB 123 TTHKLGQKYKGVGLLRLFKQAITAGRRVTRTETDGTGAVSISAAVELVHRQVDSL-- 180

QY 262 LSARMLLIGAGKMKLVKHLVAGCKKVVVNRVSRVDAIREMKDIEIVRPLSDMY 321
DB 181 -SOKTVIIGAGKMACLLVKHLLAKGATDITIVNRSQRSSQDLANQFPQAQLTLCPLTDMF 239

```

```

QY 322 QAAEADVFTSTASSETSLFAKEHAEALPPVSDTMTGGVRLFYDVISVPRNVSACVSEVGAA 381
DB 240 TAAAGDVTSTGATEPILNCENLTGCVINKSL----MLVDISVPRNVAADVHMEQV 295
QY 382 RYVNDVLLKEVVEANKEDRLKAMEAQTIITELRRFEAWRDSLETVPITIKLRSYADRI 441
DB 296 RAFNVDDDLKEVVAQNOARRQARQAEALLEEIEAIFDLWMSRLETVPITISSRSKVEDI 355
QY 442 RASELEKLCQKVGEDALTKMRRATJEELSTGVNKLHGLHQLHLCDCGSDSRTLDTEL 501
DB 356 REQELEKALSRLGSE-FAEKHGVIEALTRGVNKLHGLHPEVQLRA----QODIEARKQC 410
QY 502 MHALNRMFSLDMEX 515
DB 411 LRSKMLFDEVEE 424

RESULT 11
HEMI SYNEL STANDARD; PRT; 426 AA.
AC Q8DI53;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR).
GN HEMA OR TLL1738.
OS Synchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=32046;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shampo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
CC -|- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA(Glu).
CC -|- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC Involved in chlorophyll biosynthesis.
CC -|- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AF005375; BAC09290.1; -.
DB HAMAP; MF_00087; -.
DR InterPro; IPR000343; GluTR.
DR InterPro; IPR000594; Thif domain.
DR Pfam; PF00745; GluTR dimer; 1.
DR Pfam; PF05201; GluTR N; 1.
DR Pfam; PF05200; GluTR NAD_bind; 1.
DR TIGRFAMs; TIGR01035; hema; 1.
DR PROSITE; PS00747; GLUTR; FALSE NEG.
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
FT ACT_SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 99 99 BASE (BY SIMILARITY).
SQ SEQUENCE 426 AA; 47596 MW; D84CE5A1D2AA777E CRC64;

Query Match 34.3%; Score 893.5; DB 1; Length 426;
Best Local Similarity 44.2%; Pred. No. 4.9e-42;
Matches 193; Conservative 87; Mismatches 134; Indels 23; Gaps 5;

```

```

QY 82 IAVIGLSVHTAPVEMREKLAVAEELWPRAIQELTSLNHIEEAVALSTCNRMIEIYVVALSW 141
DB 3 IAVIGLSKHTAPVDVREKLSVPEDYRERALQHLQCYAHIQEATILSTCNRLIYIVTSDT 62
QY 142 NRGIREVVDWMSKSGIPASBLREHLFILRSSDATRHLEFVSAGLDSLVLGEGQILAQVK 201
DB 63 EVGVREVHQFLSEWHSIPLPQLRPYLFILLHQDAVMHLMRVASGLDSLIVIGEQILSQVK 122
QY 202 QVVRSGQSGGLKXNIDRMFKDAITAGKRVSETHISSGAVSVSSAAVELALMKLPKSEA 261
DB 123 RCHQLGQQYKSGIPGLNRFVFTGATAAGKRVRTETSIGTGAIVSISAAVELADRLQLNQ- 181
QY 262 LSARMLLIGAGKMGKLVIKHVAKCKKVVVNVRSVERVDAIREMKDIEIVVRPLSDMY 321
DB 182 -NCRIAVVGAGKMSLVVQHILARVGVKIRIINRSLERAQELAQOFPPEVRFELFTMTDL 240
QY 322 QAAEADVFTSTASSETSLFAKEHAEALPPVSDTMTGGVRL-----FVDISVPRNVSACV 375
DB 241 PIVAMDVLFTSTAATEPLDOR-----DNLGAVLVGDRSLAIDISVPRNVHANV 290
QY 376 SEVGARVYVDDDLKEVVEANKEDRLKAMEAQTIITELRRFEAWRDSLETVPITIKUR 435
DB 291 TELGTVQLFNVDLQAVVAQNOEARQLAQEAGILEBELETFLAMWHALETVPITIRSUR 350
QY 436 SYADKIRASELEKLCQKVGEDALTKMRRATJEELSTGVNKLHGLHQLHLCDCGSDSRTL 495
DB 351 QKMEAIRTOELEKALSRLGSE-FADKHGVIEAMTRTIINKILHDPVQLQ-----SORDL 405
QY 496 DETLENMHALNRMFSLD 512
DB 406 ESRQRAMOTLQDLFNLE 422

RESULT 12
HEMI CYAPA
ID HEMI CYAPA STANDARD; PRT; 432 AA.
AC P48077;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR).
GN HEMA.
OS Cyanophora paradoxa.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RC Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RC Bryant D.A.;
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RT Plant Mol. Biol. Rep. 13:327-332(1995).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RC Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RC Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RC Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
RN 13
RP SEQUENCE OF 407-432 FROM N.A.
RC STRAIN=UTEX LB 555 / Pringsheim;
RC MEDLINE=90092562; PubMed=2403527;
RA Bryant D.A., Stirewalt V.L.;
RT "The cyanelle genome of Cyanophora paradoxa encodes ribosomal
RT proteins not encoded by the chloroplast genomes of higher plants.";
RB PNAS Lett. 259:273-280(1990).

```

```

CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA (Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA (Glu).
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC Involved in chlorophyll biosynthesis.
CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U30821; AA811194.1; -.
CC EMBL; X17063; -. NOT_ANNOTATED_CDS.
CC F1R; T06851; T06851.
CC HSSP; Q42843; 1B29.
CC HAMAP; MF_00087; -. 1.
CC InterPro; IPR000343; GlutR.
CC Pfam; PF00745; GlutR_dimer; 1.
CC Pfam; PF05201; GlutR_N; 1.
CC Pfam; PF05200; GlutR_NAD_bind; 1.
CC TIGRfam; TIGR01035; hema; 1.
CC PROSITE; PS00747; GLUTR; FALSE NEG.
CC Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
CC NADP; Cyanelle.
CC ACT_SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
CC ACT_SITE 99 99 BASE (BY SIMILARITY).
CC SEQUENCE 432 AA; 48542 MW; F14378FC72FFD05B CRC64;
CC
CC Query Match 33.5%; Score 872.5; DB 1; Length 432;
CC Best Local Similarity 43.5%; Pred. No. 7e-41;
CC Matches 189; Conservative 86; Mismatches 148; Indels 11; Gaps 4;
CC
CC QY 82 IAVIGLSVHTAPVEMREKLAVALAEELWPRAIQELTSLNHIIEAAVLSTCNRMIEIYVALSW 141
CC DB 3 IIVVGLSHKTPVDPREKLSIPKVRIGEAIRELCNYPHIEVAIILSTCNRLIEIYLLTSDT 62
CC QY 142 NRGIREVVDWMSKSGIPASSELREHLFILRSSDATRHLFEYSAGLSLVLEGQILAQV 201
CC DB 63 YQGIREATQPLADSDLSLELURQHLFILLHQDAVMHLMRVTAGLSLIIIGEGQILSQVK 122
CC QY 202 QVVRSGQSGGLGNIDRMFKDAITAGKRVSRSETNISSGAVSVSSAAVELALMKLPKSEA 261
CC DB 123 QCYQLGQOYQIGPVNLNIFQAISAGKRVETQISTGAVSISSAAVELAQIK--KQDL 180
CC QY 262 LSARMLLIGAGKMGKLVILKLVAKCKKVVVNVRSVERVDAREMKDIEIYVRPLSDMY 321
CC DB 181 RTANITILGAGKMGRLVQHLLSKRVKDVINVRSVERAKLLVDQFKEANININLSLSELK 240
CC QY 322 QAAAEEADVFTSTASSETSLFAKEHAELPPVSDTMGGVRLFDVLSVPRNVSAACVSEVGAA 381
CC DB 241 TILQNSDIVFTGTSQSEPIIPE-----LINCDELPSLEMLFDIAVPRNVDPNVSDPNI 296
CC QY 382 RVYNVDLLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETPTIKKLSYADRI 441
CC DB 297 KVFNVVDLLKVVVSNQOQTRKMAKAAEILLEEELSAFNWGSLEAFTINKLEKAEI 356
CC QY 442 RASLEKCLKVGEDALTKMRRRAIEELSTGVNKLHGPIQLHRCGSDSRITDETLEN 501
CC DB 357 RVKLEKAIKSLRGEFVSDH-QEIVESITRGIVNKLHDPVQLRA-----QODIEIRGA 411
CC QY 502 MHALNRMFSLDMEK 515
CC DB 412 LKILQTLFNLDTIK 425
CC
CC RESULT 13
CC HEMI_VIBVU
CC ID HEMI_VIBVU
CC AC Q8DF8;
CC DT 15-SEP-2003 (Rel. 42, Created)

```

```

DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.1-) (Glutr).
GN HEMA OR VVI0254.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA (Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA (Glu).
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE016797; AA008790.1; -.
CC HAMAP; MF_00087; -. 1.
CC InterPro; IPR000343; GlutR.
CC Pfam; PF00745; GlutR_dimer; 1.
CC Pfam; PF05201; GlutR_N; 1.
CC Pfam; PF05200; GlutR_NAD_bind; 1.
CC TIGRfam; TIGR01035; hema; 1.
CC PROSITE; PS00747; GLUTR; 1.
CC Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.
CC ACT_SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
CC ACT_SITE 97 97 BASE (BY SIMILARITY).
CC SEQUENCE 419 AA; 46057 MW; 834113C720656CBD CRC64;
CC
CC Query Match 24.8%; Score 646; DB 1; Length 419;
CC Best Local Similarity 36.6%; Pred. No. 1.5e-28;
CC Matches 154; Conservative 81; Mismatches 144; Indels 42; Gaps 9;
CC
CC QY 81 TIAVIGLSVHTAPVEMREKLAVALAEELWPRAIQELTSLNHIIEAAVLSTCNRMIEIYVALS 140
CC DB 2 SLIAIGINHTASVDLREKVAFGPKLANALQLSQHEAVNGSVILSTCNRTVEYCDVKS 61
CC QY 141 WNRGIREVVDWMSKSGIPASSELREHLFILRSSDATRHLFEYSAGLSLVLEGQILAQV 200
CC DB 62 GARS--KVIDWLSQFHQIGLEELKPSLYVVEEQAAIRHLMRVSCGLSLVLGEPQILGQV 119
CC QY 201 QVVRSGQSGGLGNIDRMFKDAITAGKRVSRSETNISSGAVSVSSAAVELA---LMLKP 257
CC DB 120 QKAYSDESQOQVNDASLEKLFQKTFVAKRVTRTIDIGNRSVAVYAACTLAKHIFESLE 179
CC QY 258 KSEALSARMLLIGAGKMGKLVILKLVAKCKKVVVNVRSVERVDAREMKDIEIYVRPL 317
CC DB 180 KS-----TVLLVAGAGETIELVAKHLHANGCKMKIVANRTREALTAEQF-DAQVI--SL 231
CC QY 318 SDMYQAAAEADVFTSTAS-----ETSLFAKEHAELPPVSDTMGGVRLFDVLSV 367
CC DB 232 QEIPNHLAKADIVISSTASPLPIIGKMWETALKQRHQP-----MLLVDAIV 279
CC QY 368 PRNVSAACVSEVGAARVYNVDLLKEVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLET 427
CC DB 280 PRDVEAQVQGLNDAYLYTVDDLSQIIDSNIQKRVKVEAIQAEIYVAEESAFAFMSWLSQA 339
CC QY 428 VPTIKKLSYADRIASELEKCLKQV--GEDALTCKMRRRAIEELSTGVNKLHGPIQLH 485
CC DB 340 VDSIREYRQSANIREDLISKALLSLESGSDP-----EKLRELSNRLTNKLIHAPTRAL 394

```



```

QY 486 R 486
Db 395 Q 395

RESULT 14
HEM1_HELMO STANDARD; PRT; 443 AA.
AC Q9ZG6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamy1-tRNA reductase (EC 1.2.1.-) (Glutr).
GN HEMA.
OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OC NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99061957; PubMed=9843979;
RA Xiong J., Inoue K., Bauer C.E.;
RT "Tracking molecular evolution of photosynthesis by characterization of
a major photosynthesis gene cluster from Helicobacillus mobilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
semialdehyde + NADP(+) + tRNA(Glu).
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF080002; AAC84013.1; -
DR PIR; T31441; T31441.
DR HSP; Q42843; 1B29.
DR HAMAP; MF_00087; -; 1.
DR InterPro; IPR000343; Glutr.
DR Pfam; PF00745; Glutr dimer; 1.
DR Pfam; PF05201; Glutr N; 1.
DR Pfam; PF05200; Glutr_NAD bind; 1.
DR TIGRFams; TIGR01035; hemA; 1.
DR PROSITE; PS00747; GLUTR; 1.
KW Porphyrin biosynthesis; Oxidoreductase; NADP.
FT ACT SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
FT ACT SITE 99 99 BASE (BY SIMILARITY).
SQ SEQUENCE 443 AA; 49577 MW; E95C30E251A0C5F6 CRC64;

Query Match 24.7%; Score 643.5; DB 1; Length 443;
Best Local Similarity 34.4%; Pred. No. 2.3e-28;
Matches 157; Conservative 98; Mismatches 177; Indels 25; Gaps 9;

QY 82 IAVIGLSVHTAPVENREKLAVALAEELWPAIOELTSLNHEEAVALSTCNMEIYVVALSW 141
Db 3 IFVVLGNHKSAPVEVREKLSFTEAQLSEALHKLQMGAGIEGCCIUSTNCRTEIYGASTDM 62
QY 142 NRGIREVVDWMSKSGIPASELREHLFILRSDATRHFLFEVSAGLDSLVLGEGQILAQVK 201
Db 63 EKGMTAVKRFVLEMLQQLQPDPSKYFYVHTLYDAIRHLFRVAGSLDSNVLGSETQLGQVR 122
QY 202 QVTRSGQSGGLGKNIIDRMFKDAITAGKRVSETNISSGAVSVSSAAVELLMLPKSEA 261
Db 123 TAYORSNCNDCSGVNTWTFQQAIVTGVGRVTRTETGIDQHPVSIYSTAVELAEQVLGLKG 182
QY 262 LSARMLLITGAGWGKLVTKHLVAKGCKVWVNVRSVERVDAIREMKDIEIVYRPLSPMY 321
Db 183 RTA--MVLGAGKMSVLTUKHLVAGVGDKIIITANRSVKAEEELAKSCGGEATSF---ADVN 237

```

```

QY 322 QAAAEADVVTSTASSTSLFAKEHAERALPPVSDTWGGVRL--FVDISVPENVSACVSEVGA 380
Db 238 HRLSEADILISCTAAHTVIRKSMVE---QVMDRRGGKPVFPFDIAVRDIDPEVAQVPG 294
QY 381 ARVYNVDLKEVVEANKEDRLFKAMEAQTIITEELRRFEAMRDSLETPTIKKLSYADR 440
Db 295 THLYDIDAMQHVIDRNLAEERRKCAAEABIIIEHEINQFWRWLSLFPVFTIIVGLKNGNQ 354
QY 441 TRASELEKCLKVKGEDALTKQRRRAIEBLSTQIVNKLHGHPLOHLR--CDGSDSRTLDET 498
Db 355 IKEKELDRALCKLKH--LSEKEKCLVGLSSIVNQLLHDPITQLRHYAASPEGHLYSEI 412
QY 499 LENMHALNRMPSLDW----EKALIEQKIKAKVEKTON 531
Db 413 LQN-----LFCLDVPQGRQKRVVH--YPAVEQRQN 441

RESULT 15
HEM1_NEIMB STANDARD; PRT; 415 AA.
AC P56994;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (Glutr).
GN HEMA OR NMB0576.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OC NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAINS=WC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58.";
RL Science 287:1809-1815(2000).
CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
semialdehyde + NADP(+) + tRNA(Glu).
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB002413; AAF41004.1; -
DR PIR; C81183; C81183.
DR HSP; Q42843; 1B29.
DR TIGR; NMB0576; -; 1.
DR HAMAP; MF_00087; -; 1.
DR InterPro; IPR000343; Glutr.
DR Pfam; PF00745; Glutr dimer; 1.
DR Pfam; PF05201; Glutr N; 1.
DR Pfam; PF05200; Glutr_NAD bind; 1.
DR TIGRFams; TIGR01035; hemA; 1.
DR PROSITE; PS00747; GLUTR; 1.
KW Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.
FT ACT SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
FT ACT SITE 94 94 BASE (BY SIMILARITY).
SQ SEQUENCE 415 AA; 45480 MW; 6C5DD04A0D6B03894 CRC64;

```



```
Query Match      24.5%; Score 637.5; DB 1; Length 415;
Best Local Similarity 34.3%; Pred. No. 4.4e-28;
Matches 149; Conservative 96; Mismatches 163; Indels 27; Gaps 9;

QY 82 IAVIGLSVHTAPVEMREKLAVAEELWPRAIQELTSLNHIEEAAVLSTCNRMIEIYVVALSW 141
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3 LTAVGLNHQTAPLSIREKLAFAAALPKAVRNLRASNAATEAVILSTCNRTELYCVGDS- 61

QY 142 NRGIREVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSLVLGEGQILAQVK 201
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 ----EETIRMLADYHSLPTEERIPYLYALDMQETVRHAFRVACGLDSMWLGEPIILGQIK 117

QY 202 QVVRSGNSGGILGKNIDRMFKDAITAGKRVRSSETNISSGAVSVSSAAVELALMKLPKSEA 261
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 DAVRVAQEOESMGKKLNALFQKTFVAKVTRDTAVGENSVSASVKLAEQIPPDIGD 177

QY 262 LSARMLLIGAGKXGKLVIKHLVAKGCKVYVNVRSVERVDAIREEMKD-IEIVYRP--LS 318
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 LN--VLFIGAGEMIELVATYFAAKSPRLMTVANRTLARA---QELCDKLGYNABECLLS 231

QY 319 DMVQAAAADVVFTSTASSETSLFAKEHAALPPVSDTMGGVRLF-VDISVPRNVSACVSE 377
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
232 DLPAILHDYDVVVVSSSTASQPIVGKGMVERALKQOSM---PLFMLDLAVPRDIEAEVGD 288

QY 378 VGAARVYVNDLKEVVVEANKEDRLRKAMEAQTIITEELRRFEAWRDSLETPTIKKLSY 437
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
289 LNDAILYTYDDWNVIVQSGKEARQKAAAAAETLVSEKVAEFVRQQQGRQSVPLIKALRDE 348

QY 438 ADRIASELEKCLQKVGEDALTKQMRRAIEELSTGIVNKLHGPLQHLRCDSGSDSRTLDE 497
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
349 GEKARKQVLENAMKQAKGATAEV---LERLSVOLTNKLLHSPQTTLNKAGEEDKDL-- 403

QY 498 TLENMHALNRMFSLD 512
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
404 ----VHAVAQIYHLD 414
```

Search completed: December 15, 2003, 12:55:14  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 12:53:43 ; Search time 21 Seconds

(without alignments)

2431.694 Million cell updates/sec

Title: US-10-018-902-4

Perfect score: 2606

Sequence: 1 MATTSATTAATAAAATTAAPK.....DMEKATIEQKIKAVKTON 531

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76: \*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2313	88.8	528	T05732	probable glutamyl-
2	2225	85.4	535	T04402	probable glutamyl-
3	2113	81.1	465	T05734	glutamyl-tRNA redu
4	1784.5	68.5	552	T10186	glutamyl-tRNA redu
5	1770.5	67.9	543	E96616	hypothetical prote
6	1680	64.5	530	S65773	glutamyl-tRNA redu
7	1677	64.4	530	G86233	hypothetical prote
8	1573	60.4	542	T10245	glutamyl-tRNA redu
9	1334	51.2	524	D84718	probable glutamyl
10	910.5	34.9	463	S77180	glutamyl-tRNA redu
11	909.5	34.9	428	AG1936	glutamyl-tRNA redu
12	872.5	33.5	432	T06851	glutamyl-tRNA redu
13	643.5	24.7	443	T31441	probable glutamyl
14	637.5	24.5	415	C81183	glutamyl-tRNA redu
15	628.5	24.1	415	A81920	glutamyl-tRNA redu
16	620	23.8	419	C82109	glutamyl-tRNA redu
17	616.5	23.7	459	H84030	glutamyl-tRNA redu
18	577.5	22.2	420	AB0246	glutamyl-tRNA redu
19	573.5	22.0	418	BVECHA	glutamyl-tRNA redu
20	573.5	22.0	418	C90843	glutamyl-tRNA redu
21	573	22.0	455	A35252	glutamyl-tRNA redu
22	572.5	22.0	418	H85700	hypothetical prote
23	569.5	21.9	418	EVEBHA	glutamyl-tRNA redu
24	569.5	21.9	418	AC0720	glutamyl-tRNA redu
25	548.5	21.0	422	C83063	glutamyl-tRNA redu
26	544.5	20.9	432	A82533	glutamyl-tRNA redu
27	538.5	20.7	435	AG1431	glutamyl-tRNA redu
28	533	20.5	406	G70410	glutamyl-tRNA redu
29	523.5	20.1	448	F89950	glutamyl-tRNA redu

30	506	19.4	435	2	AB1269	glutamyl-tRNA redu
31	505.5	19.4	350	2	F75259	glutamyl-tRNA redu
32	459	17.6	468	2	H70746	probable hema prot
33	450	17.3	392	2	H64317	glutamyl-tRNA redu
34	446.5	17.1	398	1	S51136	glutamyl-tRNA redu
35	444.5	17.1	437	2	F69496	glutamyl-tRNA redu
36	435.5	16.7	402	2	A69002	glutamyl-tRNA redu
37	430	16.5	404	2	T45026	glutamyl-tRNA redu
38	428.5	16.4	449	2	G64549	glutamyl-tRNA redu
39	423.5	16.3	436	2	B84329	glutamyl-tRNA redu
40	419.5	16.1	581	2	T36267	probable glutamyl-
41	417	16.0	467	2	S72907	glutamyl-tRNA redu
42	415.5	15.9	449	2	C71959	glutamyl-tRNA redu
43	400	15.3	432	2	F81400	glutamyl-tRNA redu
44	385	14.8	416	2	D72456	probable glutamyl-
45	369	14.2	400	2	T43856	glutamyl-tRNA redu

ALIGNMENTS

RESULT 1

T05732

probable glutamyl-tRNA reductase (EC 1.2.1.1.-) 1 precursor, chloroplast - barley  
C:Species: Hordeum vulgare (barley)  
C:Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 16-Jul-1999  
C:Accession: T05732

R:Bougri, O.; Grimm, B.

Plant J. 9, 867-878, 1996

A:Title: Members of a low-copy number gene family encoding glutamyl-tRNA reductase are

A:Reference number: 215440; MUID:96267576; PMID:8696365

A:Accession: T05732

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-528 <BOU>

A:Cross-references: EMBL:X86101; NID:g1039331; PIDN:CAA60054.1; PID:g1039332

A:Experimental source: cv. Klages

C:Genetics:

A:Gene: hema1

A:Map position: 5

A:Genome: nuclear

C:Superfamily: Glutamyl-tRNA reductase

C:Keywords: aminolevulinate biosynthesis; chloroplast; NADP; oxidoreductase; porphyrin  
F:1-44/Domain: transit peptide (chloroplast) #status predicted <NP>  
F:45-528/Product: Glutamyl-tRNA reductase 1 #status predicted <MAT>

Query Match 88.8%; Score 2313; DB 2; Length 528;  
Best Local Similarity 88.1%; Pred. No. 8.6e-122;  
Matches 468; Conservative 27; Mismatches 32; Indels 4; Gaps 2;

Qy 1 MATTSATTAATAAAATTAAPKPRGSSALCORVAGGRRRSGVRCDAAGVEAQAQAAKAA 60

Db 2 MAGATSA-TAAAGAFAAAKARGPAAACPMVLVAGGRRRSGVRCDAAG---DAQAASKAA 57

Qy 61 SVAALQFKISADRYMKERSTIAVIGLSVHTAPVREKLAVALAEELWPAIQLTSLNHI 120

Db 58 SITALEQFKISADRYMKERSSIAVIGLSVHTAPVREKLAVALAEELWPAIQLTSLNHI 117

Qy 121 EEAALSTCNRMIEIYVVALSWNRGIREVDWMKSGIPASELREHLFILRSSDATHRLF 180

Db 118 EEAALSTCNRMIEIYVVALSWNRGIREVDWMKSGIPASELREHLFILRSSDATHRLF 177

Qy 181 EYSAGLDSLVLGSGQILAQVKVVRSGQNSGGIGKQIDRMFKDAITAGKVRSETNLSG 240

Db 178 EYSAGLDSLVLGSGQILAQVKVVRSGQNSGGIGKQIDRMFKDAITAGKVRSETNLSG 237

Qy 241 AVSVSSAAVELAMKLPKSEALSARMLLTAGDMGKLVIKHLVAKGCKKVVVNRVSERV 300

Db 238 AVSVSSAAVELAMKLPKSEALSARMLLTAGDMGKLVIKHLVAKGCKKVVVNRVSERV 297

Qy 301 DAIREMKDIEIYVRPLSDMYQAAAEADVFTTASSETSIPAKEHAELPPVSDTWTGGVR 360

Db 298 DAIREMKDIEIYVRPLSDMYQAAAEADVFTTASSETSIPAKEHAELPPVSDTWTGGVR 357

QY 361 LFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTITTEELRRPEA 420  
DB 358 LFVDISVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRVKAMEAQTITTEELRRPEA 417  
QY 421 WRDLSLETPTTKKLSYADRTASLEKCLQKVGEDALTKKWRRAIEELSTGIVNKLHG 480  
DB 418 WRDLSLETPTTKKLSYADRTASLEKCLQKVGEDALTKKWRRAIEELSTGIVNKLHG 477  
QY 481 PLQHLRCDSRSLTDLTLENMHALNRMFSLDMEXAIIIEQIKAKVEKTON 531  
DB 478 PLQHLRCDSRSLTDLTLENMHALNRMFSLDMEXAIIIEQIKAKVEKTON 528  
RESULT 2  
T04402  
probable glutamyl-tRNA reductase (EC 1.2.1.-) - barley  
C:Species: Hordeum vulgare (barley)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 20-Jun-2000  
C:Accession: T04402  
R:Tanaka, R.; Yoshida, K.; Nakayashiki, T.; Teuji, H.; Inokuchi, H.; Okada, K.; Tanaka, Photosyn. Res. 53, 161-171, 1997  
A:Title: The third member of the hema gene family encoding glutamyl-tRNA reductase is present in Arabidopsis thaliana  
A:Reference number: Z15335  
A:Accession: T04402  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-535 <TAN>  
A:Cross-references: EMBL:D88383; PIDN:BAAZ5169.1  
A:Experimental source: subsp. bonus  
C:Genetics:  
A:Gene: hema3  
A:Genome: nuclear  
C:Function:  
A:Description: catalyzes the reduction of glutamyl-tRNA (Glu) by NADPH to glutamic acid 1  
A:Pathway: aminolevulinate biosynthesis; porphyrin biosynthesis  
C:Superfamily: glutamyl-tRNA reductase  
C:Keywords: aminolevulinate biosynthesis; chloroplast; NADP; oxidoreductase; porphyrin b

Query Match 85.4%; Score 2225; DB 2; Length 535;  
Best Local Similarity 84.0%; Pred. No. 7.1e-117;  
Matches 452; Conservative 38; Mismatches 38; Indels 10; Gaps 5;  
QY 1 MATTTTATTAATAATTATTA---KPRGSSSALCORVAGGRRR--SGVVRCDAGVEAQAQA 55  
DB 1 MASTSTAGATAGAGAAAGVGNKPRG--SAACPRVPAGGRQKLSGVVRCD--AGQVPAQM 57  
QY 56 VAKAASVAALAEQFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALAEELWPRAIQELT 115  
DB 58 AMAASVAALAEQFKISADRYMKERSTIAVIGLSIHTAPVEMREKLAVALAEELWPRAVELT 117  
QY 116 SLNHIEEAVALSTCNRMETIYVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDA 175  
DB 118 NLNHIEKAVALSPCNRMETIYVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDA 177  
QY 176 TRHLFEVSGLDLVLGEGQILAOVQVVRSGNSGGIGKNDIRMFKDAITAGKVRSET 235  
DB 178 TRHLFEVSGLDLVLGEGQILAOVQVVRSGNSGGIGKNDIRMFKDAITAGKVRSET 237  
QY 236 NISSGAVSVSSAAVELAMKLPKSECLARMLLIGAGKMGKLVKHLVAKGCKVYVYVNR 295  
DB 238 NISSGAVSVSSAAVELAMKLPKSECLARMLLIGAGKMGKLVKHLVAKGCKVYVYVNR 297  
QY 296 SVERVDAIREEMKDIYVVRPLSDMYQAAADVVFTSTASSETSLFAKEHAELPPVSDT 355  
DB 298 SVERVDAIREEMQGGIEIYVRSITEMYEAADADVFTSTASESPLFTKEHAELPPVSGA 357  
QY 356 MGGVRLFDVDSVPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTITTEEL 415  
DB 358 LGGVRLFDVDSVPRNVSACVSDVGHARVYVNDLKEVVEANKEDRLRKAMEAQTITTEEL 417  
QY 416 RRFENRWSLETPTTKKLSYADRTASLEKCLQKVGEDALTKKWRRAIEELSTGIVN 475  
DB 418 RRFENRWSLETPTTKKLSYADRTASLEKCLQKVGEDALTKKWRRAIEELSTGIVN 477

DB 418 RRFENRWSLETPTTKKLSYADRTASLEKCLQKVGEDALTKKWRRAIEELSTGIVN 477  
QY 476 KLLHGPQLHRLCDGSDSRSLTDLTLENMHALNRMFSLDMEXAIIIEQIKAKVE--KTON 531  
DB 478 RLLHGPQLHRLCDGSDSRSLTDLTLENMHALNRMFGLDTEKAVMEQIKIKVEKQKTON 535  
RESULT 3  
T05734  
glutamyl-tRNA reductase (EC 1.2.1.-) 2 - barley (fragment)  
C:Species: Hordeum vulgare (barley)  
C>Date: 09-Apr-1999 #sequence\_revision 09-Apr-1999 #text\_change 21-Jul-2000  
C:Accession: T05734  
R:Bougri, O.; Grimm, B. Plant J. 9, 867-878, 1996  
A:Title: Members of a low-copy number gene family encoding glutamyl-tRNA reductase are present in Arabidopsis thaliana  
A:Reference number: Z15440; MUID:96267576; PMID:8696365  
A:Accession: T05734  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-465 <BOU>  
A:Cross-references: EMBL:X86102; NID:g1666078; PIDN:CAA60055.1; PID:g1666079  
A:Experimental source: cv. Bonus  
C:Genetics:  
A:Gene: hema2  
A:Map position: 5  
A:Genome: nuclear  
C:Superfamily: glutamyl-tRNA reductase  
C:Keywords: aminolevulinate biosynthesis; chloroplast; NADP; oxidoreductase; porphyrin

Query Match 81.1%; Score 2113; DB 2; Length 465;  
Best Local Similarity 89.9%; Pred. No. 1e-110;  
Matches 418; Conservative 29; Mismatches 18; Indels 0; Gaps 0;  
QY 67 QFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALAEELWPRAIQELTSLNHIEEAVAL 126  
DB 1 QFKISADRYMKERSTIAVIGLSVHTAPVEMREKLAVALAEELWPRAISELTSLNHIEEAVAL 60  
QY 127 STCNRMETIYVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDATHLFEVSAGL 186  
DB 61 STCNRMETIYVVALSWNRGIREVVDWMSKSGIPASELREHLFILRSSDATHLFEVSAGL 120  
QY 187 DSLVLGEGQILAOVQVVRSGNSGGIGKNDIRMFKDAITAGKVRSETNISAGAVSVSS 246  
DB 121 DSLVLGEGQILAOVQVVRSGNSGGIGKNDIRMFKDAITAGKVRSETNISAGAVSVSS 180  
QY 247 AAVALAMKLPKSECLARMLLIGAGKMGKLVKHLVAKGCKVYVYVNRVERVDAIREE 306  
DB 181 AAVALAMKLPKSECLARMLLIGAGKMGKLVKHLVAKGCKVYVYVNRVERVDAIREE 240  
QY 307 MKDIEIYVRPLSDMYQAAADVVFTSTASSETSLFAKEHAELPPVSDTMGGVRLFDVDS 366  
DB 241 MKDIEIYVRPLSDMYQAAADVVFTSTASSETSLFAKEHAELPPVSDTMGGVRLFDVDS 300  
QY 367 VPRNVSACVSEVGAARVYVNDLKEVVEANKEDRLRKAMEAQTITTEELRFEAWRDSLE 426  
DB 301 VPRNVSACVSEVGAARVYVNDLKEVVEANKEDRVKAMEAQTITTEELRFEAWRDSLE 360  
QY 427 TVPTTKKLSYADRTASLEKCLQKVGEDALTKKWRRAIEELSTGIVNKLHGPLOHLR 486  
DB 361 TVPTTKKLSYADRTASLEKCLQKVGEDALTKKWRRAIEELSTGIVNKLHGPLOHLR 420  
QY 487 CDGSDSRSLTDLTLENMHALNRMFSLDMEXAIIIEQIKAKVEKTON 531  
DB 421 CDGSDSRSLTDLTLENMHALNRMFNLDEKAVLEQIKIRAKVEKTON 465  
RESULT 4  
T10186  
glutamyl-tRNA reductase (EC 1.2.1.-) - cucumber  
C:Species: Cucumis sativus (cucumber)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T10186

[illegible]

## C:Genetics:

A:Gene: HBMA2  
 A:Genome: nuclear  
 A:Introns: 80/1; 160/1  
 C:Superfamily: glutamyl-tRNA reductase  
 C:Keywords: aminolevulinate biosynthesis; chloroplast; NADP; oxidoreductase; porphyrin h  
 F:1-64/Domain: transit peptide (chloroplast) #status predicted <TNP>  
 F:65-530/Product: glutamyl-tRNA reductase 2 #status predicted <MTP>

Query Match 64.5%; Score 1680; DB 2; Length 530;  
 Best Local Similarity 67.2%; Pred. No. 1.8e-86;  
 Matches 350; Conservative 60; Mismatches 85; Indels 26; Gaps 9;

```

QY 20 PRGSSALCORVAG-----GRRRSQV---RCDAGVEAQAQAAVAALEQFKI 70
DB 23 PTYSSPAPLDVIGIRALPMNNRKRGLIQRARCEISPSN-----KAASISALEQLKT 75

QY 71 SA-DRYMKERSTIAVIGLSVHTAPVEMREKLAVALAEELPRAIOELTSLNHIEEAVLSTC 129
DB 76 SAIDRYTKERSIVVIGLSIHTAPVEMREKLAIPAEWPRAIAELCGLNHIEEAVLSTC 135

QY 130 NMIEIYVALSNRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSL 189
DB 136 NMIEIYVALSNRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSL 195

QY 190 VLGEQILAQVKQVRSQNSGGLGNIDRMFKDAITAGKRVRSSTNITSSGAVSVSSAAV 249
DB 196 VLGEQILAQVKQVRSQNSGGLGNIDRMFKDAITAGKRVRSSTNITSSGAVSVSSAAV 255

QY 250 ELALMKLPK-SEALSARMLLIGAGKMGKLVIKHLVAKGCKVNVNRSVERVDATREEM- 307
DB 256 ELALMKLPK-SEALSARMLLIGAGKMGKLVIKHLVAKGCKVNVNRSVERVDATREEM- 315

QY 308 KDIEIYRPLSDMYQAABADVFTSTASSETSLFAKEHAELPPVSDTMGGVRLFPVDISV 367
DB 316 PGVEIYKPLDEMLSCAAEADVFTSTASSETPLFKEQVETLPPVRD-----ARLFPVDISV 371

QY 368 PRNVSAQVSEVGAARVNVVDLKEVVEANKEDRLRKAMEAQTIIIEELRFEAWRDSLET 427
DB 372 PRNVSAQVSEVGAARVNVVDLKEVVEANKEDRLRKAMEAQTIIIEELRFEAWRDSLET 431

QY 428 VPTIKLRSYADRIRASELEKLOKVGEDALTKQMRRAIEELSTGIVNKLHGPLOHLRC 487
DB 432 VPTIKLRSYADRIRASELEKLOKVGEDALTKQMRRAIEELSTGIVNKLHGPLOHLRC 490

QY 488 DGSRTLTDELNNHMLNMFSLDMEKAIIEQKIKAKVEK 528
DB 491 DGNDSRTLTSETLDNMQALNRMVGLDAE--ILEEKIRAKVEK 529

```

## RESULT 7

hypothetical protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 24-Aug-2001  
 R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 Jensen, N.F.; Hughes, B.; Huijzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: AB6141; MUID:21016719; PMID:11130712  
 A:Accession: G86233  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-530 <STO>  
 A:Cross-references: GB:AE005172; NID:G2160186; PIDN:AAB60749.1; GSPDB:GN00141  
 C:Genetics:

A:Map position: 1  
 C:Superfamily: glutamyl-tRNA reductase

Query Match 64.4%; Score 1677; DB 2; Length 530;  
 Best Local Similarity 67.0%; Pred. No. 2.6e-86;  
 Matches 349; Conservative 61; Mismatches 85; Indels 26; Gaps 9;

```

QY 20 PRGSSALCORVAG-----GRRRSQV---RCDAGVEAQAQAAVAALEQFKI 70
DB 23 PTYSSPAPLDVIGIRALPMNNRKRGLIQRARCEISPSN-----KAASISALEQLKT 75

QY 71 SA-DRYMKERSTIAVIGLSVHTAPVEMREKLAVALAEELPRAIOELTSLNHIEEAVLSTC 129
DB 76 SAIDRYTKERSIVVIGLSIHTAPVEMREKLAIPAEWPRAIAELCGLNHIEEAVLSTC 135

QY 130 NMIEIYVALSNRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSL 189
DB 136 NMIEIYVALSNRGIREVVDWMSKSGIPASELREHLFILRSSDATRHLFEVSAGLDSL 195

QY 190 VLGEQILAQVKQVRSQNSGGLGNIDRMFKDAITAGKRVRSSTNITSSGAVSVSSAAV 249
DB 196 VLGEQILAQVKQVRSQNSGGLGNIDRMFKDAITAGKRVRSSTNITSSGAVSVSSAAV 255

QY 250 ELALMKLPK-SEALSARMLLIGAGKMGKLVIKHLVAKGCKVNVNRSVERVDATREEM- 307
DB 256 ELALMKLPK-SEALSARMLLIGAGKMGKLVIKHLVAKGCKVNVNRSVERVDATREEM- 315

QY 308 KDIEIYRPLSDMYQAABADVFTSTASSETSLFAKEHAELPPVSDTMGGVRLFPVDISV 367
DB 316 PGVEIYKPLDEMLSCAAEADVFTSTASSETPLFKEQVETLPPVRD-----ARLFPVDISV 371

QY 368 PRNVSAQVSEVGAARVNVVDLKEVVEANKEDRLRKAMEAQTIIIEELRFEAWRDSLET 427
DB 372 PRNVSAQVSEVGAARVNVVDLKEVVEANKEDRLRKAMEAQTIIIEELRFEAWRDSLET 431

QY 428 VPTIKLRSYADRIRASELEKLOKVGEDALTKQMRRAIEELSTGIVNKLHGPLOHLRC 487
DB 432 VPTIKLRSYADRIRASELEKLOKVGEDALTKQMRRAIEELSTGIVNKLHGPLOHLRC 490

QY 488 DGSRTLTDELNNHMLNMFSLDMEKAIIEQKIKAKVEK 528
DB 491 DGNDSRTLTSETLDNMQALNRMVGLDAE--ILEEKIRAKVEK 529

```

## RESULT 8

glutamyl-tRNA reductase (EC 1.2.1.-) 2 - cucumber  
 C:Species: Cucumis sativus (cucumber)  
 C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
 R:Tanaka, R.; Yoshida, K.; Nakayashiki, T.; Masuda, T.; Taji, H.; Inokuchi, H.; Tanai,  
 Plant Physiol. 110, 1223-1230, 1996  
 A:Title: Differential expression of two hema mRNAs encoding glutamyl-tRNA reductase p  
 A:Reference number: Z16983; MUID:97088717; PMID:8934625  
 A:Accession: T10245  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-542 <TAN>  
 A:Cross-references: EMBL:D67088; NID:G1015318; PIDN:BAAL1091.1; PID:G1015319  
 A:Experimental source: cv. Aonagajibal; greening cotyledons  
 C:Genetics:  
 A:Gene: hema2  
 A:Genome: nuclear  
 C:Function:  
 A:Description: catalyzes the reduction of glutamyl-tRNA(Glu) by NADPH to glutamic aci  
 A:Pathway: aminolevulinate biosynthesis; porphyrin biosynthesis  
 C:Superfamily: glutamyl-tRNA reductase  
 C:Keywords: aminolevulinate biosynthesis; chloroplast; NADP; oxidoreductase; porphyrin

Query Match 60.4%; Score 1573; DB 2; Length 542;  
 Best Local Similarity 66.6%; Pred. No. 1.7e-80;  
 Matches 313; Conservative 74; Mismatches 79; Indels 4; Gaps 4;

Db 217 ISTLFKANKAGKRVRPAQTNIASGVSSAAVELATLTKLPQSVS--SANMLLVIGAGEMGK 275

Qy 277 LVIKHLVAKGCKVVVNRSVERVDIREMKD-IEIVYRPLSDMYQAIAEADVVFTSTA 335  
:|::||::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Db 276 RIIEHLVAKGCTKVVMNSEDKVAIRKEMSGVEIIVKPLDEILACAAEANVIFTSTS 335  
:|::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Qy 336 SETSLFAKEHAELPPVSDTMGGVRLFDVISVPRNVSAVCSEVGGAARVYNVDDLKEVREA 395  
||::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Db 336 SETPLFLKEHVEILPCPADY--ALFVDISVPRNVSCVAELDSARVYNVDDLKEVAAA 393  
||::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Qy 396 NKEDRLRKAMEAQTIITEBLRFPAWRDSLTVPTIKLRISVADRIRASELEKCLKQVGE 455  
||::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Db 394 NKEDARKSMEALPIIRETIEFGWRDSLQTFPIRKLRKTERIRAECVKLSKG- 452  
||::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Qy 456 DALTKQRRAIELSTGIYNNKLHGLOHLRGDSRDTLDETLENMHALRMFLDMEK 515  
:|::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Db 453 NGMDKKTREAVEKQTRIIYNNILDYPMKHLYDGTGSSKLRETLENMQAVNRITYELDGE- 511  
:|::||::||::||::||::||::||::||::||::||::||::||::||::||:  
Qy 516 ALIEQKIKAKVEK 528  
:||::||::||::||::||::||::||::||::||::||::||::||::||:  
Db 512 -LLEEKIREKKOK 523  
:||::||::||::||::||::||::||::||::||::||::||::||::||:

RESULT 10

S77180

glutamyl-tRNA reductase (EC 1.2.1.-) - Synechocystis sp. (strain PCC 6803)

N/Alternate names: hema protein; protein slr1808; transfer RNA-Gln reductase

C/Species: Synechocystis sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence revision 25-Apr-1997 #text\_change 16-Jun-2000

C/Accession: S77180 | A38087 | S37660

R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, J.  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, K.; Okumura, S. J. Mol. Biol. 1996, 261, 1-13

DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S77180

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-463 <KAN>

A/Cross-references: EMBL:D90308; GB:AB001339; NID:g1652725; PIDN:BAA17738.1; PID:g165151

A/Note: protein slr1808

A/Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

R/Verkamp, E., Jahn, M., Jahn, D.; Kumar, A.M.; Soell, D.

A:Title: Identification of a hema gene from *Synechocystis* by complementation of an E.  
A:Reference number: S37660; MUID:93093989; PMID:1459859  
A:Accession: S37660  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 37-463 <GRI>  
A:Cross-references: EMBL:X65963; NID:G288421; PIDN:CAA46779.1; PID:G288422  
C:Genetics:  
A:Gene: hema  
A:Start codon: GTG  
C:Function:  
A:Description: catalyzes the reduction of glutamyl-tRNA(Glu) by NADPH to glutamic acid  
A:Pathway: aminolevulinate biosynthesis; porphyrin biosynthesis  
C:Superfamily: glutamyl-tRNA reductase  
C:Keywords: aminolevulinate biosynthesis; NADP; oxidoreductase; porphyrin biosynthesis

Query Match 34.9%; Score 910.5; DB 1; Length 463;  
Best Local Similarity 43.0%; Pred. No. 1,3e-43;  
Matches 200; Conservative 85; Mismatches 163; Indels 17; Gaps 5.

Qy	82	I	A	V	I	G	L	S	V	H	T	A	P	V	N	R	E	K	L	A	V	E	L	P	R	A	I	O	E	L	T	S	L	N	H	I	E	A	A	V	L	S	T	C	N	R	M	E	I	V	V	A	L	S	141	
Db	3	I	A	V	G	L	S	H	K	T	A	P	V	E	I	R	E	K	L	S	I	P	E	O	T	S	A	I	Q	L	T	S	P	H	I	D	E	V	A	L	S	T	C	N	R	L	E	I	V	A	G	E	62			
Qy	142	N	R	G	I	E	V	D	M	S	K	S	G	I	P	A	S	E	L	R	H	L	F	I	L	S	S	D	N	T	R	H	L	F	V	S	A	G	L	S	L	V	L	G	G	O	I	L	A	Q	V	K	201			
Db	63	D	H	G	I	E	V	T	O	F	L	S	E	H	S	K	L	P	V	H	S	R	O	H	L	F	L	L	H	E	D	A	V	M	H	E	M	V	A	A	G	L	S	L	V	L	G	G	O	I	L	A	Q	V	K	122
Qy	202	Q	V	V	R	G	S	G	L	G	K	N	I	D	M	F	K	D	A	I	T	A	G	R	V	R	E	T	N	I	S	S	G	A	V	S	S	A	A	V	E	L	M	K	L	P	K	S	E	A	261					
Db	123	N	T	H	K	L	G	O	Q	O	Y	G	I	K	T	I	L	N	L	F	K	O	A	L	T	A	G	R	V	T	E	T	S	I	G	T	G	A	V	S	I	S	S	A	A	V	E	L	A	Q	I	---KAEN	179			
Qy	262	L	S	A	-	R	M	L	L	I	G	A	K	M	G	K	L	V	I	K	L	V	A	K	G	C	K	V	V	V	N	R	S	V	E	R	V	D	A	I	R	E	M	K	O	I	E	T	V	R	P	L	S	M	320	

[illegible]

Db 181 RTANITILGAGKMSRLLVQHLLSKRVKIDINIVRSVERAKLLVDQFKEANINIYNLSLK 240

QY 322 QAAAEADVFTSTASSETSLFAKEHAELPPVSDTGMGVRFLFVDISVPRNVSAACVSEVCA 381

Db 241 TILQNSDILVFTGTSQSEPIITPE-----LINCNDLPSLEMLDFIAVPRNVDPNVSPQONI 296

QY 382 RVYVDDLKVEVANKEDRLKAKMAEQIITTEELRRFEAWRDSLETPTTIKLRYSADRI 441

Db 297 KVFNVDDLKVVSQNOQTRKMAKAAABILLEELSFAFNWGSLEAITINKLREKABII 356

QY 442 RASELEKCLKVGEDALTKMRAIEELSTGIYNKLLHGPLOHLRCDGSDSRTLDLEN 501

Db 357 RVKELEKAIISLGNFVSDH-QEIVESLTRGIVNKILHDPVQLRA-----QODIEIRGA 411

QY 502 MHALNRMPSLOWEK 515

Db 412 LKILQTLFNLDTIK 425

RESULT 13

T31441

probable glutamyl-tRNA reductase (EC 1.2.1.1-) hema - Helicobacillus mobilis

C:Species: Helicobacillus mobilis

C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 15-Sep-2000

C:Accession: T31441

R:Xiong, J.; Inoue, K.; Bauer, C.E.

Proc. Natl. Acad. Sci. U.S.A. 95, 14851-14856, 1998

A:Title: Tracking molecular evolution of photosynthesis by characterization of a major P

A:Reference number: Z21036; MUID:99061957; PMID:9843979

C:Accession: T31441

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-443 <XIO>

A:Cross-references: EMBL:AF080002; NID:G3820536; PID:G3820540; PIDN:AAC04013.1

C:Genetics: hema

A:Gene: hema

C:Superfamily: glutamyl-tRNA reductase

C:Keywords: oxidoreductase

Query Match 24.7%; Score 643.5; DB 2; Length 443;  
Best Local Similarity 34.4%; Pred. No. 9.6e-29;  
Matches 157; Conservative 98; Mismatches 177; Indels 25; Gaps 9;

QY 82 IAVIGLSVHTAPVEMREKLAELWPAIQLTSLNHIIEAAVLSTCNRMIEIYVVALSW 141

Db 3 IFVVLGNHKSAPVEMREKLSFTEAQLSEALHKLQMGAGIECCILSTCNRTIELYQASIDM 62

QY 142 NRGIREVVDWMKSGIPASERLHFLIRSSDATHLFEVSAGLDSLVLGEQILAQVK 201

Db 63 EKGMTAVKRFVLEWQLOPQDFSKFYVHTLYDAIRHLFRVASGLDSMVLGETQLGQVR 122

QY 202 QVRSQNSGGLGNIDRMFKDAITAGKRVSETNISGAVSVSAAVELALMKLPKSEA 261

Db 123 TAYQRSNEDCSNGIINTFWFOALTVGKRVKRTETGIDQHPVSIYSYAVELAEQVLGLKG 182

QY 262 LSARMLLIGAGKMGKLVIKHLVAKGCKVNVNRSVERVDAIREMKDIELTVRPLSDMY 321

Db 183 RTA--MVLGAGKMSVLTIKHLVAGVGDKIITANSVEKAEELAKSCGGEALSF--ADVN 237

QY 322 QAAAEADVFTSTASSETSLFAKEHAELPPVSDTGMGVRFLFVDISVPRNVSAACVSEVGA 380

Db 238 HRLEADILISCTAATHYVIRKSMVE---QVMDRRGKPKVPFFDIAVPRDIDPEVAQVPG 294

QY 381 ARVNVVDDLKVEVANKEDRLKAKMAEQIITTEELRRFEAWRDSLETPTTIKLRYSADR 440

Db 295 THLYDIDAMQHVIRDNLAERKCAEABIIIEHINOFMRWINSLFVIPTIVGLKNKGQ 354

QY 441 IRASELEKCLKVGEDALTKMRAIEELSTGIYNKLLHGPLOHLR--CDGSDSRTLDLET 498

Db 355 IKEKELDRALCKLKH--LSEKEKLVGLASSIYNQLLHDPITQLRHYAASPEGHLYSEI 412

QY 499 LENHNLNRMPSLDM-----EKAIIEQKIKAKVEKTON 531

Db 413 LQN-----LFCLDVPQQRKHVVVH--YPAVEQRQN 441

## RESULT 14

C81183

glutamyl-tRNA reductase NMB0576 [imported] - Neisseria meningitidis (strain MC58 sero

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: C81183

R:Pettersen, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: C81183

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-415 <TET>

A:Cross-references: GB:AE002413; GB:AE002098; NID:G7225797; PIDN:AAF41004.1; PID:G722:

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB0576

C:Superfamily: glutamyl-tRNA reductase

Query Match 24.5%; Score 637.5; DB 2; Length 415;  
Best Local Similarity 34.3%; Pred. No. 1.9e-28;  
Matches 149; Conservative 96; Mismatches 163; Indels 27; Gaps 9;

QY 82 IAVIGLSVHTAPVEMREKLAELWPAIQLTSLNHIIEAAVLSTCNRMIEIYVVALSW 141

Db 3 LTAVGLNHQTAPLSIREKLAFAAALPAVNRNLARSNAATEAVILSTCNRTELYCVGDS- 61

QY 142 NRGIREVVDWMKSGIPASERLHFLIRSSDATHLFEVSAGLDSLVLGEQILAQVK 201

Db 62 ----EEIIRWLADYHSLPIEBIRPYLYALDMQETVRHAFRVACGLDSMVLGEQILGQIK 117

QY 202 QVRSQNSGGLGNIDRMFKDAITAGKRVSETNISGAVSVSAAVELALMKLPKSEA 261

Db 118 DAVRAVOEQESMGKLNALFOKTFEVAKEVTRTDTAVGNSVSMASASVKLAIEQIPDGD 177

QY 262 LSARMLLIGAGKMGKLVIKHLVAKGCKVNVNRSVERVDAIREMKD-IEIVRYP--LS 318

Db 178 LN--VLFIGAGEMIELVATYPAKSPRLMTVYANRTLARA----QELCDKLGVAEPCLLS 231

QY 319 DMYQAAAEADVFTSTASSETSLFAKEHAELPPVSDTGMGVRFLF-VDISVPRNVSAACVSE 377

Db 232 DLPAILHDYDVVVSSTASQLPIVCKGMVERALKQRQSM--PLFMLDLAVPRDIEAEVGD 288

QY 378 VGAARVNVDDLKVEVANKEDRLKAKMAEQIITTEELRRFEAWRDSLETPTTIKLRSY 437

Db 289 LNDAYLTVDDMVNIVQSGKEARQKAAAAAETLVSEKVAEFVROOQGRQSVPLIKALRDE 348

QY 438 ADIRASELEKCLKVGEDALTKMRAIEELSTGIYNKLLHGPLOHLRCDGSDSRTLDE 497

Db 349 GEKARKQVLENAMKQLKAGTAEV---LERLSVQLTNKLLHSPTQTLNKAGEEDKDL-- 403

QY 498 TLENHNLNRMPSLSD 512

Db 404 ----VHVAQIYHLD 414

## RESULT 15

A81920

glutamyl-tRNA reductase (EC 1.2.1.1-) NMA0760 [imported] - Neisseria meningitidis (str

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: A81920

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo;

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.



5028 22 : 21177 000